Insight-driven Crisis Information - Preparing for the Unexpected using Big Data

Hendrik Stange
Fraunhofer IAIS
hendrik.stange@iais.fraunhofer.de

Sylvia Steenhoek
BBK
sylvia.steenhoek@bbk.bund.de

Sebastian Bothe
Fraunhofer IAIS
sebastian.bothe@iais.fraunhofer.de

François Schnitzler
Technion
francois@ee.technion.ac.il

ABSTRACT

National information and situation centers are faced with rising information needs and the question of how to prepare for unexpected situations. One promising development is the access to vastly growing data produced by distributed sensors and a socially networked society. Current emergency information systems are limited in the amount of complex data they can process and interpret in real-time and provide only partially integrated prediction and alarming capabilities. In this paper we present a novel approach to build a new type of automated and scalable information systems that intelligently make use of massive streams of structured and unstructured data and incorporate human feedback for automated incident detection and learning. Big data technologies, uncertainty handling and privacy-by-design are employed to match end-user system requirements. We share first experiences analyzing data from the centennial flood in Germany 2013.

Keywords

Big Data, Reality Monitoring, incident detection, uncertainty, crisis information.

INTRODUCTION

Digitalization is a game changer in crisis management. However, the systematic integration of social media into emergency information systems in Germany is still in their infancy (Zisgen & Kern, 2014). Aggravating this circumstance, nation-wide natural disasters and extreme weather events have recently seen unforeseen magnitudes and unexpected social responses. Kern and Zisgen recognized a behavioral change when analyzing Facebook in the aftermath of the centennial flood in Germany, 2013. They expect Facebook or other social media to have an increased relevance for crisis management of the future (Kern & Zisgen, 2014).

Other studies revealed that the lack of public communication and missing official information resulted in volunteered information mapping using Web 2.0 tools (Breuer, 2014). Instantaneous access to information affects the expectations of the public in times of crisis. But there is not a single source of information for sharing information and coordinating help. This results in redundancy of information on different systems holding the risk of conflicting facts. Without the ability to know what is being shared and planned via social channels there will always be a blind spot in professional Emergency Information Systems (EIS).

There is also a risk of sharing false or outdated information. This is one reason that hinders authorities like the Federal Office for Civil Protection and Disaster Assistance (BBK) to provide secured information in near real-time as they have to manually analyze and verify information from data they receive. Breuer claimed that the public expects faster information provisioning and uses crowdsourcing themselves as a mean to collect and validate crisis information (Breuer, 2014). Thus, new methods to validate the trustworthiness and fast collection of relevant data get in the focus of professionals.

Consequently, a considerable number of works has dealt with the design of crisis information systems (Turoff, Chumer, van de Walle, & Xao, 2004). Common objective was on creating situation awareness and knowledge management as well as support of messaging, assessment, and resources planning by human experts (deNIS, 2011). With the instrumentation of our world with diverse connected sensors, Smartphones, and social networks massive volumes of data describing our urban environment become available offering the potential of enhanced services and automated support of detection and monitoring of crisis situations.

In this paper we propose a novel approach for building insight-driven Emergency Information Systems (iEIS) that bridges the divide between data science and disaster management. Our goal is to develop a data system for joint information and situation centers that makes use of big data technologies and that is agnostic in terms of data sources to detect critical incidents under uncertainty.

RELATED WORK

With the proliferation of social networks the utilization for crisis management and communication has been investigated (Litou, Boutsis, & Kalogeraki, 2014; Klafft, 2013) as well as its effects on disaster response (Kern & Zisgen, 2014). Other work has spanned the area of reality monitoring working on engines to detect emergency events early in advance and build comprehensive, dynamic situation maps based on collaborative crisis management (Stange & Bothe, 2013; Purohit, Meier, Castillo, & Sheth, 2013). A goal of reality monitoring is to provide personalized information services and privacy-aware warning and prediction using massive data streams of various sensors.

Processing and analyzing massive volumes of data has received a lot of attention in the research area as well as in the digital industry. Companies like Google, Twitter, and Amazon invested to build new software frameworks and tools to support storage, processing and querying vast amounts of data from the web, users and systems. Nathan Marz and James Warren proposed a meta-concept named *Lambda Architecture* for designing big data systems (Marz & Warren, 2013). They envisioned a modularized decomposition of a big data application (storage, processing, result view, and application) and divided the data processing into two layers: *batch layer* to perform operations on all data and *speed layer* to apply in-stream functions as the data enters the system. Finally the results of both layers are merged and presented to the application. An introduction to the core concepts on designing big data applications is given in (Mock, Sylla, & Hecker, 2014). One of the keys to build big-data-ready information systems lies in its *horizontal scalability*, that is, by adding new hardware nodes, limitations in data storage or processing power can be resolved leaving the application itself untouched. EIS even require to extend this scalability to data sources that feed into the system.

However, these concepts and developed technologies have not yet been fully used for building Emergency Information systems. Secondly, the accessibility of hundreds of complex data sources (e.g. social data sources like Twitter, Instagram, Whatsapp, Vimeo, Youtube) make it infeasible to handle them on an individual level or to provide only partial, single-source information to the end-user.

Vital element in disaster management is knowledge. In order for data systems to operate we need a structured knowledge base. Many work has been done Knowledge Management Systems. Please refer to (Dorasamya, Ramanb, & Kaliannan, 2013) for a comprehensive overview. Structured knowledge, e.g. as linked data, extends data-based Emergency Information Systems by providing means to manage derived insights and build on top of an EIS.

PRACTICAL SYSTEM REQUIREMENTS

In the following we will review requirements that have been identified together with the City of Dublin City Council (Dermot Kinane) and the Federal Office for Civil Protection and Disaster Assistance in Germany (BBK, Julia Kern). To ensure interoperability and transferability we included city-level and nation-wide perspectives. From a professional perspective we aim at next generation EIS that build on top of data systems to be scalable, flexible, fault tolerant, secure, trustable, collaborative, relevant, and prepared for automation.

Scalability has two aspects: One is the range of the EIS scaling from city-level to nation-wide foci; and the second aspect comes from a system design perspective and means horizontal scalability which states that the complexity of the underlying big data system is obscured from the application through a standard system interface. Simultaneously, this realizes resource-savings and cost-efficiency constraints as it allows to use commodity server hardware and start with a smaller sized computer cluster (easy extensibility).

Flexibility refers to the system ability to include and understand new data sources as they become available and secondly, to incorporate new types of crisis events as they become known or by requisite of new end-users of the system. The system should be decoupled from the data producers (e.g. sensors) and independent of the techniques for data dissemination (e.g. via API, as RSS feed or website).

Fault tolerance refers to failsafe operations of the system in case of hardware failure and the unavailability of one or many input data sources due to internal hardware failure or lost connection to the data provider. The system should continuously monitor its health and trustworthiness and report to the user.

Security of the system addresses compliance issues such as privacy which should be taken into account from the beginning. Other security concerns relate to the accessibility of the data by a human operator, data retention period, and data fusion regulations.

Trustable EIS ensure a high degree of certainty in all information provided and visualized to the end-user. It holds mechanisms to measure the trustworthiness of information, data sources and derived findings. Transparency of decisions is realized by a tracing mechanism that logs the event detection process.

Collaborative features of EIS embrace a participatory approach for the automated management of resources and to improve emergency detection and validation of detected events as well as data enhancement in smart cities and countries.

An overall goal is **process automation** from data collection up to information mapping, prediction and alarming. This non-exclusively comprise:

- automated detection of relevant events and alarming (awareness)
- automatically crowdsource selected human users for targeted measurements/ feedback
- dispatch semi-automatic responses
- provide labels and descriptions of events (including confidence level)
- seamlessly integrate human experts in the decision making process

Relevance of information becomes a corner stone as more and more data enters the situation centers. EIS should present information in a transparent and usable way building their own situation understanding. Preferences of end-users should also be taken into account.

With respect to the scope of this paper, please note that we do not explicitly mention common usability and visualization requirements. Our focus, however, is on sustainable system requirements that emerge from the data revolution and new expectations by professionals.

REFERENCE SYSTEM DESIGN

More data, unexpected crisis events, and limitations in user perception require novel insight-driven Emergency Information Systems (*iEIS*). To be prepared for the dynamics in the field of big data, machine learning/ analytics and crisis response and management, we are looking for novel system architectures that internalize all the above mentioned system requirements.

Especially, legal constraints and the expected massive production of data in a connected digital world increase the relevance for automated analysis and guided reality monitoring. We, therefore, propose a novel architecture for *iEIS* to support the situation awareness process (see Figure 1) by (1) analyzing various data sources from traffic and social media to mobile activity for insights on events, and by (2) providing a situational overview and condensing the events to a big picture.

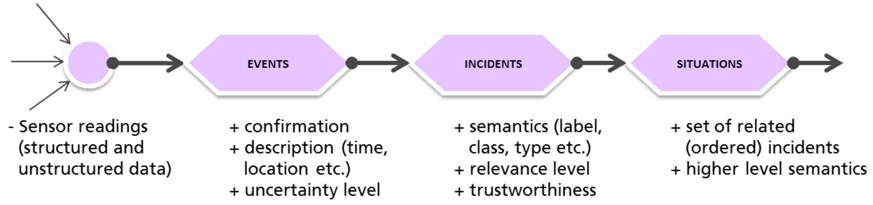

Figure 1: Common data flow and situation awareness process

The process in Figure 1 already illustrates the general workflow and functioning of an *iEIS* (see Figure 2). Massive volumes of poly-structured (sensor) data enter the distributed *iEIS* via sensor-specific access interfaces and are collected, processed and analyzed in designated Intelligent Sensor Agents (**ISA**). Each ISA uses adaptive automated event detection to recognize known events and discover anomalies in the data stream of its assigned data source. These detected event candidates are send to an information fusion and uncertainty handling unit, we call Round-Table (**RT**). At the RT other sources of information are queried to support and thereby verify the identified abnormality of the originating sensor. If enough evidence on the event is provided, it is confirmed and the system uses an event ontology to assign the event to a known incident class. This ontology may utilize Global Incident Knowledge from disaster knowledge management systems (see (Dorasamya, Ramanb, & Kaliannan, 2013)). If not enough information is provided (high uncertainty) but the significance of the identified anomaly remains, the system may use crowdsourcing to ascertain the finding and to provide a label for the event which can be used for training future event detection. Participatory means may also be used to enrich information on an event. The system provides feedback on the uncertainty of the decision and logs the decision making process to learn from historic events and provide user feedback on relevance and accuracy to be able to automatically identify known events in the future.

Detected events are stored in a database and are made available via an application server to the end-user via a flexible, multi-platform dashboard. Additional analytics capabilities are used to visualize complex situations based on the single incidents including information on situation dynamics or extension. Building on top of this process we will have emergency resource planning, situation foresight or reporting processes. Figure 2 shows the general system architecture of an *iEIS* that applies big data principles and follow participatory design paradigms. Thus, the system is transparent and permeable for user feedback and traceable to understand decision making across the system.

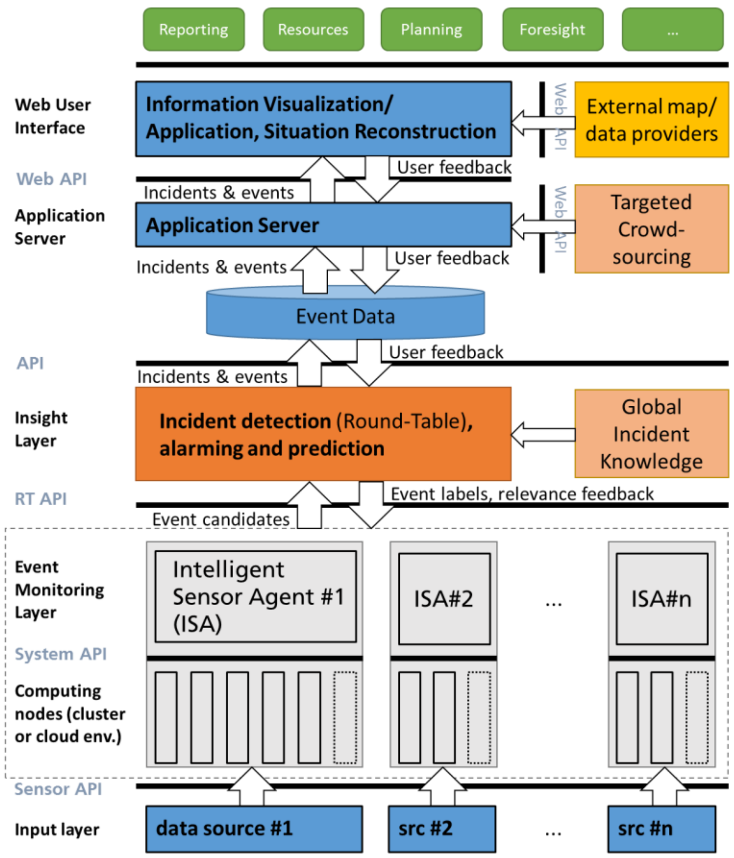

Figure 2: General system architecture of insight-driven Crisis Information Systems

Privacy and compliance

Aside transparency any *iEIS* relying on potentially personal data must comply with a number of laws and regulations. Utilizing an indefinite number of data sources, legal compliance becomes a major challenge. The proposed architecture provides a number of mechanism to integrate privacy-by-design concepts, such as, avoid fusion of raw data and decouple the sensor data processing from the event detection process. This way the ISA module may even be located at different locations or stays with the data owner. The open design of the ISA supports customized implementation for flexibility.

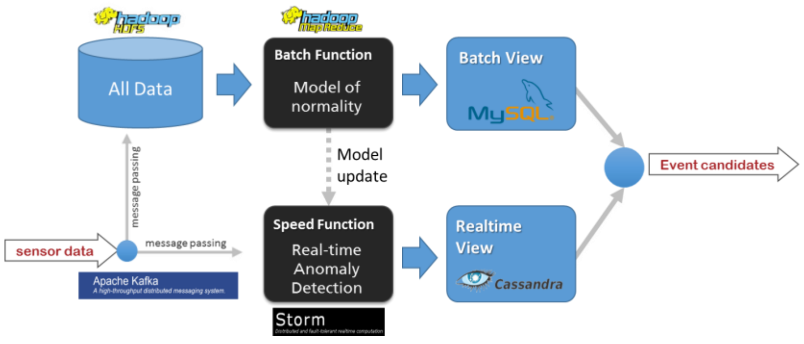

Figure 3: Implemented ISA architecture following the Lambda Architecture

Anomaly detection

The detection of potentially relevant events is performed at ISA-level at near sensor level where each ISA understands the nature of the monitored sensors and data floating into the system. Customized methods are used from areas like machine learning. Any advancements to event detection or changes at sensor level typically result in changes on a single ISA (single point-of-adaptation strategy).

To be ready for data streams, each ISA instantiates the Lambda Architecture using standard big data technologies, e.g. Storm[1] or Streams[2] for stream processing, Hadoop[3] for raw data storage and batch processing, databases such as Cassandra[4] or MySQL for storing results, see Figure 3.

The idea is to efficiently distribute two different types of data processing tasks: the batch layer stores and processes all data and typically performs building and adapting a model of normality; the speed layer handles incoming data in real-time to detect in-stream events using the normality model.

The ISA concept allows to have multiple ISA-Sensor-relationships depending on the task/ perspective on the data. For example, a Twitter-Event-ISA uses Deep Learning methods to identify emergency events on Tweets (Paass & Pratap, 2014) while other Twitter-Event-ISA look for emotional indications of crisis events (Valkanas & Gunopulos, 2013) or emergency topics (Fuchs, Stange, Samiei, Andrienko, & Andrienko, 2015).

Information fusion and uncertainty handling

Decentralization of anomaly detection enforces a centralized event-based information fusion process. Confirming events and assigning a label requires independent information sources and background knowledge to validate, enhance, and annotate event candidates. A probabilistic Round-Table-Approach has been proposed to automatically detect incidents under uncertainty (Schnitzler, et al., 2014). To communicate between RT and ISA and for semantic enrichment of events we use an event ontology. Therein hierarchical relationships between types of events are encoded (e.g. flood → natural → disaster → incident). If the confidence level in the decision is low – contradictory or too limited in support by other ISA – we propose to integrate targeted crowd-sourcing. We aim to ask people in the field specifically for reliably providing local insights or to confirm events. We deploy a system developed by the University of Athens for efficient

[1] https://storm.apache.org/
[2] http://www-ai.cs.uni-dortmund.de/SOFTWARE/streams/
[3] http://hadoop.apache.org/
[4] http://cassandra.apache.org/

event detection mechanisms and reliable crowdsourcing (Boutsis, Kalogeraki, & Gunopulos, 2013) and Ohmage[5] as a backend server.

Visual situation awareness

Typically a large number of incidents is recorded in the vicinity of a large scale emergency. On city level we have traffic congestions or local flooding, nation-wide effects exhibit crevasses, blackouts or mass evacuations. The more data sources we connect to the system the more aspects and events will be recorded. To support the work at the information center we integrate visual analytics methods and guided visualizations. In particular we use dynamic clustering of events to reconstruct situational aggregates (Andrienko, Andrienko, Fuchs, & Stange, 2014) and zoom-level dependent event clustering.

REAL WORLD TEST BED

Germany faced a serious flood in June 2013. As the water level increased, so did the importance of social media. Mainly Facebook and Twitter were used to share information and organize help. Data from Twitter and mobile networks have been used to demonstrate basic operation of the system (Fuchs, Andrienko, Andrienko, Bothe, & Stange, 2013). It became evident that each single source had its blind spots and uncertainties. Especially the resolution of the data sources indicated challenges for event validation. Additional ISA tapping new sources or perspectives are required as well as crowdsourcing to assist incident detection at the Round-Table level.

These initial tests raised the attention for such a system that allows, among others, privacy-aware social media monitoring in combination with sensor data inputs. Such a system will improve the understanding and visualization of the situation as well as estimating mood and multi-scale situational insights. But we also experiences limitations regarding the trustworthiness of information. Thus, a multi-modal approach using various data sources and crowdsourcing to validate information will gain importance.

[5] http://ohmage.org/

CONCLUSION

The increasing importance of crowd-sourcing and social media in reality monitoring and disaster response as well as the propagation of real-world sensors open up new possibilities to advance emergency information systems and to put new tools in the hands of disaster managers. In particular, a better societal management of the overall cycle of disaster monitoring and response becomes possible, citizens may become involved in emergency detection and data acquisition/ validation, and advanced planning can economize resources. On the other hand we see many novel requirements and constraints when putting big data into operations. Tapping new sources of information will help us to deal with uncertainty in single-source, real-world data and to raise awareness of unforeseen anomalies for early warning. Event ontologies structure the detection process and support communications. Big Data Analytics and machine learning methods reduce manual efforts and efficiently automatize the entire emergency detection process. What is required is a big data mind-set and new skills at the information and situation centers to open up for the new possibilities in big data which the people have already utilized for their life.

FUTURE WORK

The reference architecture described will be deployed at the BBK for assessment of operational fitness. Furthermore, together with our partners at the Academy for Crisis Management, Emergency Planning and Civil Protection we are currently developing realistic scenarios for training purposes.

ACKNOWLEDGEMENT

This work has been supported by the EU-FP7 INSIGHT project (Grant No. 318225). A substantial amount of groundwork has been done by our partners at the University of Athens, University of Dortmund and at Technion.

REFERENCES

1. Andrienko, N., Andrienko, G., Fuchs, G., & Stange, H. (2014). Detecting

and Tracking Dynamic Clusters of Spatial Events. *IEEE VAST.* Paris: IEEE.

2. Boutsis, I., Kalogeraki, V., & Gunopulos, D. (2013). Efficient Event Detection by Exploiting Crowds. *The 7th ACM International Conference on Distributed Event-Based Systems* (S. 123-134). Arlington, TX, USA: ACM.
3. Breuer, D. (2014). Dann mach' ich es eben selbst. *Civil Protection Magazine*(3), 26-29.
4. deNIS, F. O. (2011). *German Emergency Preparedness and Response Information System (deNIS IIplus)*. (BBK) Abgerufen am 28. 01 2015 von http://www.bbk.bund.de/DE/AufgabenundAusstattung/Krisenmanagement/deNIS/deNISIIplus/deNIS_zweiplus_einstieg.html
5. Dorasamya, M., Ramanb, M., & Kaliannan, M. (2013). Knowledge management systems in support of disasters management: A two decade review. *Technological Forecasting and Social Change*, 1834-1853.
6. Fuchs, G., Andrienko, N., Andrienko, G., Bothe, S., & Stange, H. (2013). Tracing the German Centennial Flood in the Stream of Tweets: First Lessons Learned. *GeoCrowd at ACM SIGSPATIAL* (S. 31-38). Orlando, FL, USA: ACM.
7. Fuchs, G., Stange, H., Samiei, A., Andrienko, G., & Andrienko, N. (2015). A semi-supervised method for topic extraction from micro postings. *Information Technology*, S. to appear.
8. Kern, J., & Zisgen, J. (2014). "I like Hochwasser". *Bevölkerungsschutz (Transl.: Civil protection)*(3), 17-19.
9. Klafft, M. (2013). Diffusion of emergency warnings via multi-channel communication systems an empirical analysis . *IEEE Eleventh International Symposium on Autonomous Decentralized Systems (ISADS).* Mexico: IEEE.
10. Litou, I., Boutsis, I., & Kalogeraki, V. (2014). Efficient Dissemination of Emergency Information using a Social Network. *Mining Urban Data (MUD) - EDBT/ICDT.* Athen, Greece: CEUR-WS.
11. Marz, N., & Warren, J. (2013). *ig Data – Principles and best practices of scalable real-time data systems.* USA: Manning Publications.
12. Mock, M., Sylla, K.-H., & Hecker, D. (2014). Skalierbarkeit und Architektur von Big-Data-Anwendungen. *OBJEKTspektrum*.
13. Paass, G., & Pratap, B. (2014). Semantic Role Labeling Using Deep Neural Networks. *The International Workshop on Representation Learning.* Nancy, France.
14. Purohit, H., Meier, P., Castillo, C., & Sheth, A. P. (2013). Crisis Mapping, Citizen Sensing and Social Media Analytics: Leveraging Citizen Roles for Crisis Response. *7th International AAAI Conference on Weblogs and Social Media*.
15. Schnitzler, F., Liebig, T., Mannor, S., Souto, G., Bothe, S., & Stange, H. (2014). Heterogeneous Stream Processing for Disaster Detection and Alarming. *IEEE International Conference on Big Data*, (S. 914-923).
16. Stange, H., & Bothe, S. (2013). Reality Monitoring - Sicherheitswerkzeuge vernetzter Gesellschaften. *Crisis Prevention*(2).
17. Turoff, M., Chumer, M., van de Walle, B., & Xao, X. (2004). The Design of a Dynamic Emergency Response Management Information System (DERMIS). *The Journal of Information Technology Theory and Application (JITTA)*, 1-35.
18. Valkanas, G., & Gunopulos, D. (2013). Event Detection from Social Media Data. *IEEE Data Eng* (S. 51-58). IEEE.
19. Zisgen, J., & Kern, J. (2014). Social Media im Bevölkerungsschutz (Transl.: Social Media in Civil Protection). *Public Security*(1), 26-25.