Towards Intelligent Disaster Response Systems

Visar Shehu
South East European University, Macedonia
Adrian Besimi
South East European University, Macedonia
Urim Vejseli
Center for Crisis Management, Macedonia
Douglas Jones
MIT Lincoln Lab, Lexington, Massachusetts, United States

Abstract

Next Generation Incident Response System (NICS) is a platform developed by MIT Lincoln Lab that is currently being used in Macedonia by the Centre for Crisis Management (CCM). It allows coordination during natural disasters between first responders of various departments and allows them to use state of the art tools to communicate and share information. This research focuses on advancing the platform by introducing intelligent agents to the platform, based on machine learning techniques and natural language processing. Our goal is to leverage data generated in social media and feed NICS with automatically processed information from these media categorized in twelve different needs (categories). This paper presents the current state of our research, preliminary results and final goals.

Introduction

The Next-Generation Incident Command System (NICS) is a platform developed by the MIT Lincoln Lab (Bremyer, 2011; Di Ciaccio et al., 2011). Its main goal is to assist first responders by allowing them a better situational awareness and provide them with tools to collaborate and communicate during natural disasters. The platform is designed to be scalable and reliable. It enables all entities involved as first responders, to use a cross platform web application to share information via diverse tools that include geospatial information, various topologies, vehicle location tracking, weather, critical infrastructure etc. In addition, the system is scalable and modular and can be extended to suit the needs of different entities that use it.

In 2016, NICS was adopted by NATO and implemented in Bosnia and Herzegovina, Croatia, Montenegro, and the Former Yugoslav Republic of Macedonia (DHS Science & Technology Press Office, 2016).

Advances in information systems, have made IT a commodity. Since internet connection is ubiquitous, it has become the communication tool of choice for most of the population. According to Takahashi et al. (2015), Twitter has become a key communication channel that complements traditional channels during national

disasters. Vieweg et al. (2010) also recognize the importance of social media during national disasters. Their research was based on Oklahoma grassfires in 2009, and how the information from social media could be used for situational awareness.

Kongthon et al. (2012), similarly used data gathered by Twitter during the Thai floods of 2011, to assist local communities in need.

One major issue when gathering information from social media is that the information will be written in a local language and will most probably be written in a slang or misspelled. These languages are also lacking NLP capabilities, often recognized as Low Resource Languages. A project that aims to tackle these problems is DARPA's LORELEI (Low Resource Languages for Emergent Incidents) Program (Strassel et al., 2016). For this research, we were given access to this platform with the purpose of leveraging the capabilities of the system and enhancing NICS with such results.

Methodology

As previously mentioned, our main goal was to develop tools that would enhance NICS and leverage from the algorithms developed by project LORELEI. For this purpose, we needed text written in languages spoken in Macedonia (Macedonian, Albanian and Turkish) collected in social media during national disasters. Besides the text of the message, we also collected relevant metadata (most importantly information about time and location).
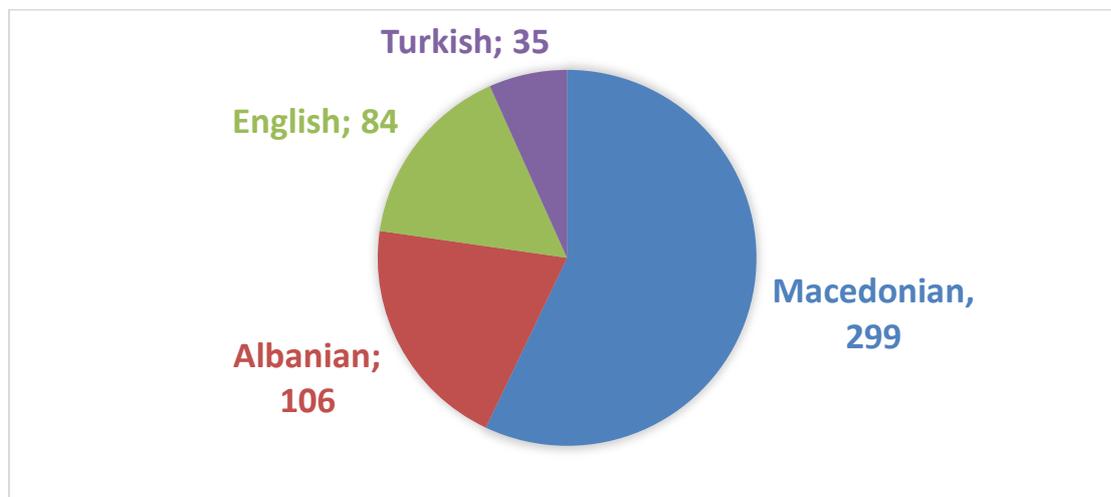
Data Gathering

We conducted two experiments, the first one was to gather data from social media and the second one was to simulate such data.

For the first experiment, two social media platforms were targeted: Facebook and Twitter. From Facebook we used the Facebook Graph API to collect all public posts from the floods in Macedonia of 2015 and 2016. We also queried twitter with geofencing data to collect tweets posted from the affected regions during the same time. All the tweets and Facebook posts gathered were treated as truth data (training set). The problem with this approach was the large amount of noise that was collected. The dataset suffered from two major problems: one was that our crawlers collected data that was not relevant to the events (e.g., plenty of normal tweets from the selected locations), and tweets that were somehow related to the event but were not actionable (e.g., people complaining or blaming the government). After pre-processing, from a total of 5613 posts during this time period, only 172 were usable.

The second experiment was a simulated event. It was conducted on a controlled environment, with a group of 33 students from South East European University. Each student was put in a situation and assigned a specific role from a set of three roles: Brave Citizen, Fearful Caregiver and Complaining Citizen. A brave citizen would be someone posting neutral, non-negative and factual posts. The second role was of someone who takes care of other people (parent, caretaker, doctor in a hospital etc) and is worried about the safety of others. Finally, the third category was one of a complaining citizen, a person who posts rumours, complains about events and does not rely his posts on facts. Every student was put in a specific situation, in a specific role. They were given enough information to post data, but we took special care not to influence their post style or what they would say. The simulation lasted three hours and allowed us to collect 524 tweets. The following chart illustrates their distribution.

Figure 1
Text Distribution by Language

Source: Authors' illustration

This second approach allowed us to gather more data and all of it was relevant to the event. Students were encouraged to use a natural language and the result was that the written text contained spelling / grammatical errors as well as usage of slang. The only drawback of this approach was that it was missing geotagged metada as well as the time of the event was not trustworthy.

Data pre-processing

Our second step was to analyse this data and extract relevant entities. We used automatic entity extraction to extract location relevant information (cities, roads, landmarks) as well as a manual approach to get better results.

Automatic entity extraction was implemented by combining a list of known entities affected regions by the flood. This way the system was able to recognize places and automatically tag them with geolocation data. This method worked for most of our data, the rest was manually analysed. Obviously, this approach was suitable due to the small scale of data we had, however it would not be feasible in a real time scenario. Despite this, when gathering data from social media, most of the posts reveal geolocation data.

Detection of needs

The LORELEI platform is designed to detect actionable information from low resource languages. Its goal is to be capable of detecting relevant information only 24 hours after an incident has occurred. Figure II shows more information about the capabilities of the platform. For our purposes, we needed to classify tweets into twelve different categories. These categories basically represented needs (in literature one might find references to needs as 'colors'), among which were: evac, food, shelter, water, utils etc.

Figure 2
The anticipated time for LORELEI to be able to understand a low resource language

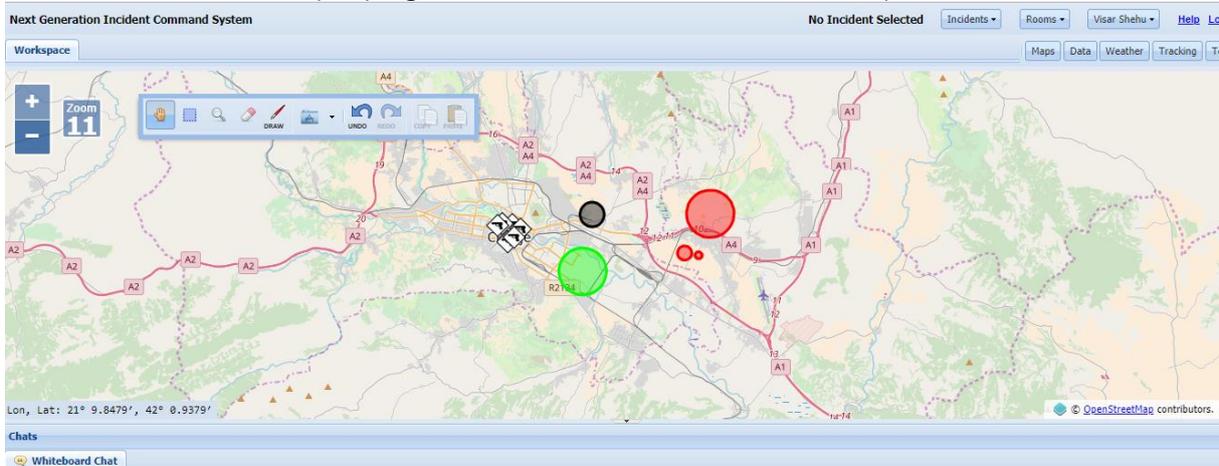|	1 Day	1 Week	1 Month
Place Names	●	●	●
Topic Spotting in Text	◗	●	●
Sentiment/Emotional State	◗	●	●
Partial Translation	◗	◖	●
Events	◗	◖	◖
Relationships	○	◖	●
Person Information	○	◖	●
Topic Spotting in Speech	○	◖	◖
Full Translation	○	◗	◖
Speech to Text	○	○	◖

Source: Templeton (2015)

When textual data were posted to the LORELEI platform, the system would respond with a Situation Frame data structure. This data structure contains the following properties: **DocumentId**, **Type** (which represents the need), **Place** (geolocation data), **Status** (current or not current), **Confidence** (ranging from 0 to 1), **Urgency** (true or false) etc.

Results

Responses from LORELEI and the social media tweets we gathered were aggregated and posted into the NICS platform. Depending on the number of tweets per location, we color coded the responses and created heatmaps in the system as in Figure III. Each color represents a given need type, and serves first responders to use the platform and react to new intelligence gathered by social media. The social media layer can be shown and hidden by demand so it does not interrupt the workflow of first responders.

Figure 3
The NICS Interface Displaying Needs as Color Coded Heatmaps

Source: Authors' illustration (from NICS interface)

Discussion

The results here show only a limited part of our research into integrating social media into the NICS platform. We limited our approach into only using social media input to evaluate the possibility of extending NICS with intelligent features.

Currently we are working on two other areas of improving the platform. At first we plan to gather all the chat between first responders that has happened inside NICS. The chat will be analysed through NLP with the purpose of better understanding the activities of first responders. The results can be further used to realize if a need has been met by responders or not.

The second step is to consider the Standard Operating Procedure of CCM (SOP) which thoroughly defines what actions should be taken and resources used to react in specific incidents. By matching chat information from inside NICS and social media input, authorities can be given automatic recommendations and use the platform as a Decision Support System.

Conclusion

This paper shows our initial steps towards extending NICS with machine learning capabilities. While the current research is limited only in detecting needs by the population affected by a natural disaster, it has a potential to be much further extended. These features will be able to assist first responders in Macedonia to act on intelligence gathered by social media.

References

1. Bremyer, P. (2011), "Next-Generation Incident Command System", Tech Notes, available at: https://www.ll.mit.edu/publications/technotes/TechNote_NICS.pdf (11 June 2018)
2. DHS Science & Technology Press Office (2016), "News Release: NATO Adapts NICS, a Communications Platform for First Responders", available at: https://www.dhs.gov/science-and-technology/news/2016/11/28/news-release-nato-adapts-nics-first-responders (11 June 2018)
3. Di Ciaccio, R., Pullen, J., Breimyer, P. (2011), "Enabling distributed command and control with standards-based geospatial collaboration", in proceedings of the 2011 IEEE International Conference on Technologies for Homeland Security (HST), 15-17 November 2011, Waltham, MA, USA, IEEE, pp. 512-517.
4. Kongthon, A., Haruechaiyasak, C., Pailai, J., Kongyoung, S. (2012) "The role of Twitter during a natural disaster: Case study of 2011 Thai Flood", in proceedings of the PICMET '12: Technology Management for Emerging Technologies, 29 July – 2 August 2012, Vancouver, BC, Canada, IEEE, pp. 2227-2232.
5. Strassel, S., Tracey, J. (2016), "LORELEI Language Packs: Data, Tools, and Resources for Technology Development in Low Resource Languages", in proceedings of 10th International Conference on Language Recourses and Evaluation (LREC), 23-29 May 2016, Portorož, Slovenia, European Language Resources Association (ELRA), pp. 3273-3280.
6. Takahashi, B., Tandoc Jr, E. C., Carmichael, C. (2015), "Communicating on Twitter during a disaster: An analysis of tweets during Typhoon Haiyan in the Philippines", Computers in Human Behavior, Vol. 50, pp. 392-398.
7. Templeton, G. (2015), "How DARPA Plans to Decrypt the Languages That Computers Still Don't Understand", available at: https://motherboard.vice.com/en_us/article/d73ezj/how-darpa-plans-to-decrypt-the-languages-that-computers-still-dont-understand (11 June 2018)
8. Vieweg, S., Hughes, A. L., Starbird, K., Palen, L. (2010), "Microblogging during two natural hazards events: what twitter may contribute to situational awareness",

in proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 10-15 April 2010, Atlanta, Georgia, USA, ACM, pp. 1079-1088.

About the authors

Visar Shehu has a PhD in computer science and is employed as an Assistant Professor at the CST department of South East European University. His main research interests are related to data mining and machine learning technologies with a focus on text data and natural language processing. He is also involved in applying his research in the industry, being the CTO of a software development company in Macedonia that develops HR solutions for the EU market. The author can be contacted at **v.shehu@seeu.edu.mk.**

Adrian Besimi has a PhD in computer science and is employed as an Assistant Professor at the CST department of South East European University. His interests are mainly related with application of IT in business and other organizations, such as: Business Information Systems, e-Commerce, Web and Mobile Solutions, Software Oriented Architectures, Data Mining and Business Intelligence and similar. The author can be contacted at **a.besimi@seeu.edu.mk.**

Urim Vejseli, PhD is the Head of Department for NATO and International Cooperation of the Crisis Management Centre of Macedonia. He is responsible for implementation of the NICS platform in the country. He also holds the position of Associated Professor at the department of Security Policies of Biznesi College in Kosovo. The author can be contacted at **urim.vejseli@cuk.gov.mk.**

Doug Jones is a member of the technical staff in the Information Systems Technology Group at MIT Lincoln Laboratory. His background includes A.B. and A.M. degrees in Linguistics from Stanford University specializing in computational phonology, and a Ph.D. in linguistics from the Massachusetts Institute of Technology specializing in Hindi syntax. He completed postdoctoral work on computational theories of verb structure at MIT's Artificial Intelligence Laboratory, and the University of Maryland's Institute for Advanced Computer Studies (UMIACS). The author can be contacted at **daj@mit.ll.edu.**