Finding Actionable Information on Social Media

Ammar Ahmed Abdelhalim
University of Cincinnati
abdelhad@mail.uc.edu

Jess Kropczynski
University of Cincinnati
jess.kropczynski@uc.edu

ABSTRACT

A great deal of data is generated every day on social media, although this information is used for marketing purposes regularly, it has the potential to serve other purposes, such as in crisis management. This study focuses on collecting data from social media, specifically Twitter, in order to help 911 telecommunicators (floor supervisors, call takers, and dispatchers) to 1) identify Twitter users requesting assistance during a crisis, 2) identify information that may be useful to incidents that were called into 911, and 3) pass the information to the first responders (police, fire, and emergency medical services). Previous research in this area can be summarized into three stages. First, a set of information requirements has been developed that must be satisfied to dispatch first responders and meet their immediate awareness needs. Second, a coding schema has been presented to identify six types of actionable information. Finally, it proposed automated methods based on previous literature which can be used to implement these methods in the future (Kropczynski et al. 2018). This research concentration is on refining social media data by starting with finding local tweets that contain this information and recognize patterns of how it is used. Next, patterns will be used in the development of automated methods in the future. The contribution of this work is extending the coding schema of the 6Ws and put it on an action, develop an interface to view the data of social media separated by the 6Ws. It will begin with just on of the six Ws (Weapons).

IT Research Symposium'19 Extended Abstracts, April 9, 2019, Cincinnati, Ohio, USA.

INTRODUCTION

Large amounts of data are posted to social media every second—this is untapped potential to recognize patterns that we didn't know existed and can have real-time application to crisis response. For example, video surveillance, it is a ubiquitous task done every day. Companies may employ an individual to monitor a small number of cameras on a corporate building that could be tracked easily by individuals and on occasion, these cameras have offered useful information about developing crises. Today, hundreds of cameras are deployed at any given time from drones to cell phones creating more information than what a human can feasibly monitor. With the help of computers using machine and deep learning methods, we are able now to identify and predict human behaviors through cameras (Kropczynski et al. 2018). This was accomplished by gathering a lot of data and automating to a recognition of specific actions to predict their probability of committing a crime.

Automated techniques have included methods to identify relevant tweets that lack obvious identifiers such as keywords or hashtags related to a crisis event (Herndon and Caragea, 2016; Li et al., 2018). Techniques have also been developed to filter bots and other irrelevant information (Wang et al., 2014; Varol et al., 2017). As researchers have recognized that a great deal of information generated about crises are shared by news sources and external observers (Olteanu et al., 2015; Starbird et al., 2010), several studies have turned their attention to identifying eyewitness accounts that may be useful to first responders (Caragea, Squicciarini, Stehle, Neppalli, and Tapia, 2014). However, despite the extensive development of techniques to identify and filter event-relevant information, the question remains, what information do emergency dispatchers and first responders really need?

The question has been answered throughout criteria comprise “golden tweet” and a coding schema “5W’s and Weapons”. To begin with, a “golden tweet” is a tweet (posted on Twitter) that contains actionable information for telecommunicators that they can pass to first responders in order to take actions. By speaking with administrators and telecommunicators they have been asked to create fictitious tweets that contain the information they are looking for. Then, they came to a better understanding of the protocol they are using to filter information from 9-1-1 calls, which can be used to identify and filter the information through social media posts. As a result, a coding schema has been initiated, "5W's + Weapon", they refer to What, Where, When, Who, Why and Weapons. By filling in information related to these 6W's, you will be able to get all the actionable information that first dispatchers need to act. Although simple to gather in a one-on-one phone call with an eyewitness, identifying all information necessary to extract actionable information from social media is quite difficult. For each of the above Ws, a special sorting and filtering technique is necessary. For this reason, rather than focusing broadly on each of the 6Ws, we will approach each individually. This research focuses specifically on identifying tweets that mention the use of a weapon.

METHODS AND TOOLS

For this research, I will identify methods to identify tweets that mention weapons in the local area and display them in a visualization too. I will use the ‘Elastic Stack’ which mainly consists of Elasticsearch, Logstash, and Kibana. To begin with, Elasticsearch is an open source, RESTful, distributed search and analytics engine built on Apache Lucene, it is commonly used for log analytics, full-text search, security intelligence, business analytics, and operational intelligence use cases. It works through sending JSON documents to Elasticsearch using API or by using Logstash. Elasticsearch automatically stores the original document and adds a searchable reference to the document in the cluster’s index. Then search and retrieve the document using the Elasticsearch API. The advantages of Elasticsearch are easy

to get started, high performance, it comes with handful tools and plugins which allow piping and visualization, reading and writing data takes less than second to complete which could be used for near real-time applications and it supports several programming languages (aws.amazon.com).

Logstash is a light-weight, open-source, server-side data processing pipeline that allows you to collect data from a variety of sources, transform it on the fly, and send it to your desired destination. It is most often used as a data pipeline for Elasticsearch (aws.amazon.com). The advantages of Logstash allows ingesting unstructured data easily from data sources like system logs, website logs, and application server logs. Next, Kibana, which is an open-source data visualization and exploration tool used for log and time-series analytics, application monitoring, and operational intelligence use cases. It offers powerful and easy-to-use features such as histograms, line graphs, pie charts, heat maps, and built-in geospatial support, it comes with geospatial capabilities so it could layer in geographical information on top of data and visualize it on maps, has a pre-built aggregations and filters which helps in analytics work and it could easily create reports of the data (aws.amazon.com).

Lists of common words used to describe shootings and gunshots will be added to filter tweets in this category and lists of common weapon types will also be used. By integrating visualization of tweets by location and filtering by the presence of weapon keywords, we hope to begin to tackle identification of one of six Ws. Future work will combine this information with other data layers to maximize the potential of this resource.

CONCLUSION

After an extensive effort, the previous work has helped to describe the type of information that emergency dispatchers and first responders really need, that happens after understanding the protocol and the techniques that telecommunicators use to extract actionable data from 9-1-1. Then, applied these techniques on social media (Twitter), which lead in the end to the coding schema 6Ws. The aim of this paper is to extend the coding schema of the 6Ws and put it on action, develop an interface to view the data of social media separated by the 6Ws, which could be extended for teaching machine learning algorithms to put it as a real-time tool after giving it a huge number of well-organized filtered data, to identify people who are really in risk during crises which will support the work of emergency dispatchers, and it could uncover the unexpected. It will begin with just one of the six Ws and will work to apply additional techniques iteratively.

REFERENCES

[1] Caragea, C. Squicciarini, A., Stehle, S., Neppalli, K., and Tapia, A. (2014) Mapping moods: Geo-mapped sentiment analysis during hurricane sandy. ISCRAM 2014 Conference Proceedings - 11th International Conference on Information Systems for Crisis Response and Management, (May), 642–651. Retrieved from http://www.iscram.org/legacy/ISCRAM2014/papers/p29.pdf

[2] Herndon, N., and Caragea, D. (2016) An evaluation of approaches for using unlabeled data with domain adaptation. Network Modeling Analysis in Health Informatics and Bioinformatics, 5, 25. https://doi.org/https://doi.org/10.1007/s13721-016-0133-6

[3] Jess Kropczynski, Rob Grace, Julien Coche, Shane Halse, Eric Obeysekare, Aurélie Montarnal, et al. (2018). Identifying Actionable Information on Social Media for Emergency Dispatch. In Kristin Stock, & Deborah Bunker (Eds.),

Proceedings of ISCRAM Asia Pacific 2018: Innovating for Resilience – 1st International Conference on Information Systems for Crisis Response and Management Asia Pacific. (pp. 428–438). Albany, Auckland, New Zealand: Massey University.

[4] Olteanu, A., Vieweg, S., and Castillo, C. (2015) What to Expect When the Unexpected Happens: Social Media Communications Across Crises. Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15, 994–1009. https://doi.org/10.1145/2675133.2675242

[5] Starbird, K., Palen, L., Hughes, A. L., and Vieweg, S. (2010) Chatter on the red: what hazards threat reveals about the social life of microblogged information. CSCW '10 Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work, 241–250. https://doi.org/10.1145/1718918.1718965

[6] Wang, A. H., Hai, A., Detecting, W., Bots, S., Networking, S., Machine, S. A., and Wang, A. H. (2014) Detecting Spam Bots in Online Social Networking Sites: A Machine Learning Approach, HAL Id: hal-01056675, 0–8.