

Finding Incident-Related Social Media Messages for Emergency Awareness

Alexander Nieuwenhuijse¹(✉), Jorn Bakker¹, and Mykola Pechenizkiy²

¹ Coosto, Fellenoord 35, 5612 Eindhoven, The Netherlands
{alexander.nieuwenhuijse,jorn.bakker}@coosto.com

² Eindhoven University of Technology,
PO. Box 513, 5600 Eindhoven, The Netherlands
m.pechenizkiy@tue.nl

Abstract. An information retrieval framework is proposed which searches for incident-related social media messages in an automated fashion. Using P2000 messages as an input for this framework and by extracting location information from text, using simple natural language processing techniques, a search for incident-related messages is conducted. A machine learned ranker is trained to create an ordering of the retrieved messages, based on their relevance. This provides an easy accessible interface for emergency response managers to aid them in their decision making process.

Keyword: Incident related social media monitoring

1 Introduction

With the ever growing social media networks, and the information shared on those networks, the opportunity for automated data analysis is increased with it. A large part of this data is generated by private users, posting about their everyday life and their surroundings. By combining the data generated by the users of these social media platforms new insights can be gathered concerning the geographic hotspots about which people talk. This information can help in the decision making process of emergency response units and their managers, who need to make decisions about sending aid to incident locations.

In the work of MacEachern et al. [3] a survey was conducted among emergency managers from International Association of Emergency Managers (IAEM) and FirstResponder.gov about their current and envisioned use of geovisual tools that support social media analysis. This survey shows that only 39.1% of the participants use social media to gather information from the public. The most important feature (94.7%) requested by the emergency managers is the inclusion of maps showing geographical information of an incident. The second most requested feature (71.1%) is the option to search through photo and video collections relating to the incident, since these graphics allow a domain expert to assess a situation.

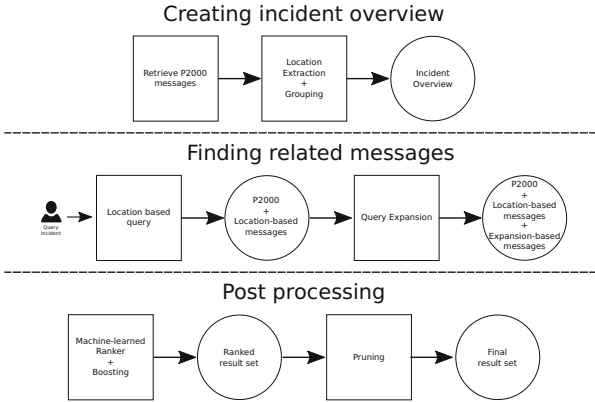


Fig. 1. Overview of the developed framework

These survey results show a relatively low use of social media data to gather information, due to the large quantity of social media messages and the time consuming process to analyse this data. As such, there is a demand for a more structured and more accessible representation of social media data. The framework described below is integrated within the Coosto solution and enables emergency response units to easily access these social media messages to aid them in their decision making process.

2 Framework

Figure 1 shows an overview of the developed framework. Only one input is required from the user: a selection of an incident for which to retrieve related social media messages. The other components of the framework operate fully automated.

An overview of active incidents is created from a set of P2000 messages. These messages are created by emergency response managers and are publicly broadcasted over the P2000 paging network to alert and dispatch emergency response units towards a reported incident. Due to the nature of these messages they are always highly reliable and contain exact location information of the reported incidents, thus creating a solid basis from which to start searching for other, non-P2000, messages.

Using the extracted location an initial search is conducted to retrieve social media messages mentioning the incident location. Even if these messages are not directly related to the incident they can be used to create a context, and also to get in contact with people at, or traveling towards, the incident location. To retrieve harder to find messages, not explicitly mentioning the incident location, a query expansion step is applied. This query expansion extracts relevant terms from the intermediate result set and uses these terms to construct a more



Fig. 2. Media objects related to the Chemelot fire. They inform about the location and scale of the incident, wind direction, shape and color of the smoke plume. (Color figure online)

complex query. This query expansion is tuned towards a high recall, to ensure that we retrieve as many relevant messages as possible.

Finally, to reduce the noise in the final result set, a machine learned ranking is applied to the results. To accomplish this we considered multiple ranking algorithms. Train and test sets were created by manually labeling messages in the result sets as either being *very relevant*, *somewhat relevant* or *irrelevant* and the performance of the rankers was measured using the Normalized Discounted Cumulative Gain [2]. From these experiments the RankBoost [1] algorithm yielded the best results and was adopted into the framework. The rank assigned by the machine learned ranker is also boosted, based on the occurrence of terms retrieved from the query expansion step. This boosting is applied to compensate for the out-of-vocabulary words in the ranking model and to cover incident types that were not present in the training set of the machine learned ranker.

Combining all these steps yields the result set of social media messages, automatically retrieved and ranked, ready to be presented to the user.

3 Usecase - Chemelot

In the morning of November 9th 2015 a large fire developed at a chemical complex called Chemelot, located near Geleen, the Netherlands. Using the developed framework the media objects shown in Fig. 2, related to this fire, were obtained. As the idiom states: “A picture is worth a thousand words”, these photos provide a lot of information for an emergency response manager, located in a response



Fig. 3. Concerned tweet asking about ammonia hazard.

center far away from the incident location. A fireman can use these photos to estimate how the fire is developing based on the smoke trail, color and direction. These initial pictures can help in the decision making process of what units to dispatch to the incident location.

As shown in Fig. 3, people are asking questions about toxicity of the fumes. These messages can be important for press officers and the related information management. Given that the fire was located at a chemical complex, and by analysing the color of the smoke, an alert can be given to the people living in the areas effected by the smoke, with the advice to stay indoors.

The overview created by the framework can easily give insights for the emergency response units and support in their decision making process. It allows for easy interaction with concerned people and informing the public, without it being a time consuming process.

4 Conclusion

The developed framework provides easy access to incident-related social media messages and helps providing emergency awareness. This application can be used by anyone. It is no longer a time consuming process to construct complex queries, and it does not require domain specific knowledge to operate. It enables emergency response managers in their decision making process, making social media more accessible.

References

1. Freund, Y., Iyer, R., Schapire, R.E., Singer, Y.: An efficient boosting algorithm for combining preferences. *J. Mach. Learn. Res.* **4**, 933–969 (2003)
2. Järvelin, K., Kekäläinen, J.: Cumulated gain-based evaluation of ir techniques. *ACM Trans. Inf. Syst.* **20**(4), 422–446 (2002)
3. MacEachren, A.M., Jaiswal, A., Robinson, A.C., Pezanowski, S., Savelyev, A., Mitra, P., Zhang, X., Blanford, J.: Senseplace2: geotwitter analytics support for situational awareness. In: 2011 IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 181–190. IEEE (2011)