

Chapter 4

Robust End-User-Driven Social Media Monitoring for Law Enforcement and Emergency Monitoring



Birgit Kirsch, Sven Giesselbach, David Knodt, and Stefan Rüping

Introduction

It has been widely proven that in case of emergencies and crisis situations content published on social media platforms, such as Twitter, constitutes a valuable source of information. As stated in Imran et al. (2015), Twitter has been used in various crisis scenarios to identify events, monitor event progress, summarize tweets to give a situational overview and classify them to extract relevant content. In this paper we propose a social media monitoring and analysis framework that provides support to the security sector and is based on the needs of two running H2020-projects in the area of community policing and emergency response, CITYCoP and E2MC.

CITYCoP, short for Citizen Interaction Technologies Yield Community Policing, aims at developing applications that facilitate community policing. One idea behind the project is to mine social media platforms to assess the perceived quality of life of the citizens.

The purpose of the **E2MC project**, short for Evolution of Emergency Copernicus services, is to extend the available Copernicus Emergency Management System (EMS) that provides an early warning and rapid mapping services in case of emergencies. By integrating an analysis platform capable of extracting crisis-related information from social media, the quality of the delivered information should be improved and the speed in which the affected area can be located and damage can be assessed should be increased. The main problem we face in both use cases is to flexibly adapt to new scenarios and provide an analysis pipeline that can be operated

B. Kirsch (✉) · S. Giesselbach · D. Knodt · S. Rüping
Fraunhofer Institute for Intelligent Analysis and Information Systems, Sankt Augustin, Germany
e-mail: birgit.kirsch@iais.fraunhofer.de; sven.giesselbach@iais.fraunhofer.de;
david.knodt@iais.fraunhofer.de; stefan.rueping@iais.fraunhofer.de

© The Author(s) 2018
G. Leventakis, M. R. Haberfeld (eds.), *Community-Oriented Policing and Technological Innovations*, SpringerBriefs in Criminology,
https://doi.org/10.1007/978-3-319-89294-8_4

by a user without expert knowledge in data analytics. In order to achieve this, we propose a framework that combines different machine learning approaches and can be enriched by user knowledge about the targeted event or location.

Related Work

A variety of methods have been applied to social media in the past in order to classify crisis-related tweets and to obtain useful information. One simple method is the keyword matching approach. In a predefined keyword dictionary, relevant words are collected. According to their appearance in a message text tweets can be extracted or classified (Olteanu et al. 2014). In more sophisticated approaches machine learning techniques are applied to develop models that are capable of automatically classifying tweets. These can be separated into supervised and unsupervised methods. Supervised methods learn models based on an already labeled dataset and can then be applied to incoming tweets. Caragea et al. (2016) and Nguyen et al. (2016) use online deep learning to classify tweets as informative or not informative. Nguyen et al. (2016) additionally partition crisis-related tweets into information-specific subclasses. Most of the frameworks proposed and developed in the context of event- and crisis-analysis, such as Tweedr (Ashktorab et al. 2014) or AIDR (Imran et al. 2014), are built upon these supervised techniques which rely on labeled datasets. We decided to integrate an unsupervised learning method that is able to train models without the need for a labeled set to overcome this drawback. To the best of our knowledge, there are only few examples in literature that follow this approach. Imran and Castillo (2015) use an unsupervised topic modeling technique, LDA, to generate candidate categories for tweet classifications. Kireyev et al. (2009) explore, how topics models can be used to analyze disaster-related Tweets. One reason why these techniques are not yet widely spread is that they uncover hidden structures that are not necessarily of interest for a user. Wang et al. (2016) present an extension of LDA, the Targeted Topic Model, that is capable of building topics around an aspect of interest. Based on this algorithm we propose an end-user driven approach to analyze tweets with respect to the location, relevant hashtags or sentiment.

System Overview

Our system contains the following components to process crawled Tweets: The **text processing component** applies typical text preprocessing techniques, such as sentence splitting, tokenization and stemming, to prepare the tweets for further analysis. The data is then enriched by integrating Hashtags, URL categories and handles as additional features. Furthermore, a user can define an event-profile with a georeference and the time the event started. Based on the profile, we calculate the

distance of the tweets geolocation to the area of interest and the time elapsed since the event started. The **sentiment analysis component** identifies the polarity of texts in terms of whether the texts are emotionally positive, negative or neutral. This so called sentiment is added as an additional feature for the topic extraction component. The **topic extraction component** is the core step of our approach and will be described in detail in section “[End-User-Driven Topic Detection](#)”. The **visualization component** is used to display enriched tweet data and enables the user to monitor tweets based on their location, derived sentiment and assigned topics. Tweets can be clustered based on their density and according to the assigned topics or keywords contained. Additionally, interactive filters based on keywords and topics can be applied to support specific scenarios.

End-User-Driven Topic Detection

Most text analytic models rely on a set of parameters or constraints that have to be set in advance of the model training. Since these parameters can heavily influence the model performance, this training process often requires complex tuning and expert data science knowledge. To allow a user to influence the analysis pipeline in a more natural way, we focus on an extension of the broadly used topic modeling algorithm LDA (Blei et al. 2003), the targeted topic model (Wang et al. 2016), that is capable of generating flexible models and integrating user domain knowledge about the event of interest.

LDA assumes that each document can be described as a mixture over topics, where each topic represents a distribution over words. A major problem one faces when applying LDA to analyze tweets is, that topics of interest may not be detected and the topics reflecting the dataset are too coarse. The idea of targeted topic modeling is to focus on an aspect the user is interested in by setting target keywords in advance. Based on this target a topic model is generated. This model consists of one major topic reflecting the content not related to the aspect and multiple fine-grained topics related to the aspect. The topics are represented as word distributions. By defining a set of target keywords related to an event of interest, these topic models can be used to extract tweets related to an event and to uncover useful structures to cluster these tweets. Figure 4.1 shows the analysis process compared to the baseline LDA-approach and a keyword based approach. In the keyword-based approach, a keyword list is used to filter relevant tweets while in the LDA approach, keyword lists are used to identify relevant topics by comparing them to the topic-word-distributions. In both scenarios, the selection of these keywords plays a crucial role and heavily affects the quality of the retrieved content. The advantage of the targeted topic model is, that it automatically extends the keywords assumed to be relevant based on the input keyword list and the word co-occurrences in the tweets. This allows a more intuitive keyword selection without the need for expert refinement. To cover a rich spectrum of relevant information, we extend the model and enrich the input data with additional keywords representing different

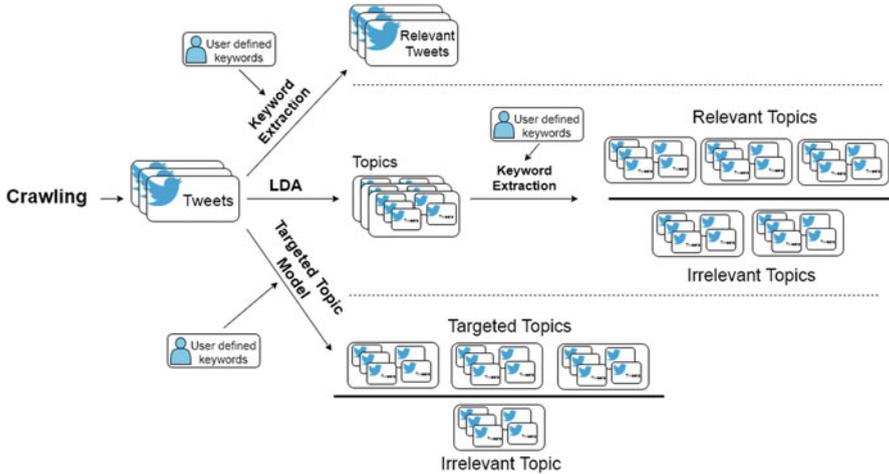


Fig. 4.1 Identification of relevant tweets

forms of information, such as the sentiment, the location and other meta data. We chose keywords as the way the information is exchanged, since it is easy to inspect and understand for the end-user and can be extended to other relevant types of information.

Experiments and Results

To evaluate our methods, we perform an experimental analysis of tweet data related to three crisis-events, the flood in Alberta in 2013, the Bombings in Boston during and after the marathon in April 2013 and the explosion at the West Fertilizer Company storage and distribution facility in April 2013. Therefore, we merged two publicly available datasets, CRISISLEX26 (Olteanu et al. 2015) and CRISISLEX6 (Olteanu et al. 2014) that were labeled from crowd-source workers as “on-topic” and “off-topic” and are assigned to one of six categories, e.g., “Infrastructure and utilities” and “Donations and volunteering”. The label information assigned to each tweet is only used to evaluate the performance of the trained model. The analysis itself is performed on the unlabeled dataset. After preparation, we generated three datasets with around 10,000 tweets each. To evaluate our analysis approach on these dataset, we simulated an interactive analysis workflow that consists of three steps: The identification of event-related tweets, the identification and analysis of topic structures and the visual analysis.

Identification of event-related tweets: based on an input keyword set a targeted topic model is trained and applied to classify tweets as relevant or not relevant. To examine the robustness of the model, we pick very general and simple target

Table 4.1 Comparison of tweet-classification techniques

Model-type	Texas explosion			Boston bombing			Alberta flood		
	Precision	Recall	FMeas.	Precision	Recall	FMeas.	Precision	Recall	FMeas.
Targeted TM	0.973	0.835	0.900	0.938	0.698	0.800	0.937	0.980	0.958
LDA	0.817	0.903	0.858	0.816	0.589	0.684	0.812	0.901	0.854
Keyword	0.995	0.796	0.885	0.959	0.660	0.782	0.957	0.976	0.977

Table 4.2 Top 20 words for the identified sub-categories

Infrastructure damage:

Bowriver, downtown, emerg, calgari, bridg, update, medicinhat, state, local, remain, citi, power, rcmp, water, school, close, canmor, break, expect, elbow

Donations:

Relief, donat, love, redcross, great, text, peopl, victim, find, made, good, feel, affect, bad, provid, send, pretti, assist, custom, fan

keywords for model training and the keyword matching. For the Alberta Flood the target keywords are set to all words matching the substring “flood”, for the explosion in Texas we pick all words matching the substring “explo” and the substrings to identify the Boston event are “bomb” and “marathon”. To assess the performance of the model, the results are compared to two different baseline-approaches:

- a keyword matching based approach, in which tweets are filtered as relevant when they contain a word in the specified keyword set.
- the direct application of LDA, where each tweet assigned to a topic identified as crisis-related is treated as relevant. We assumed all topics with high word probability for the words listed in the defined keyword set to be crisis-related.

Table 4.1 shows the performance for each approach. In all three datasets the targeted topic model outperforms the baseline LDA regarding F1 score. In the Alberta dataset, the identification seems to be not that challenging and the use of the targeted topic model approach cannot improve the performance. One reason could be that most of the tweets contain flood-related hashtags, such as #albertaflood or #abflood.

Identification and analysis of topic structures: In the next phase we further analyze the learned topic-word-distribution. In the experiment, we focused on the Alberta dataset and chose a model that consists of 25 topics describing the different hidden structures in the event-related tweets and one major topic assigned to all tweets identified as not event-related. In order to focus on a specific aspect of interest, we concentrate on the topics best matching manually defined keywords, in this case “blocked / closed / damaged” for damage related topics and “donate / fund / relief” for donation related topics. The results are presented in Table 4.2. The damage-focused topics not only contain words describing infrastructure, such as road, school or bridge, but are also capable of uncovering affected rivers, e.g. the Elbow River or affected towns, such as Medicine Hat, and damaged areas, such as the Bowness Park.

Table 4.3 Meta features with high relevance in the damage related and donation related topics

Handles		Hashtags		Sentiment	
Damage	Donation	Damage	Donation	Damage	Donation
Nenshi	redcrosscanada	#yycflood	#abflood	Neutral 67%	Neutral 65%
Calgary police	Mookalicious	#nature	#yycflood	Negative 18%	Positive 19%
Calgary herald	redcrossab	#abflood	#yychelps	Positive 15%	Negative 16 %
City of calgary	westjet	#yyc	#canmore		
Druh farrel	premier_redford	#CHLive	#medhat	Distance	
Weather network	wbrettwilson	#stampede	#mhflood	Damage	Donation
Mookalicious	nateinvegas	#flames	#heartbreaking	-100 km 100%	1000 km + 39.5%
Carrie tait	starbucks	#sunnyside	#disgusted		100 km - 28.5%
CalgaryPolice	cityofcalgary	#UCalgary	#shame		1000 km - 20.5%
globeandmail	yychelps	#staysafe	#blackhawk		5000 km + 11.5%

Next, the meta features assigned to the topics are examined. Table 4.3 lists the handles, locations, hashtags, time and URLs related to damage- or donation-topics. While in the damage-related topics the fact, that a tweet position is close to the event, is very relevant, in the donation topic this is not that important. The detected URLs as well as the handles and hashtags for the donation-topic mainly focus on institutions offering help, such as the Red Cross and politicians providing support. In the damage-related topics the focus is set to news agencies, the police and the affected towns.

Visual Analysis: To obtain a better insight, we visually examine the trained topic model. Therefore, a sentiment-map is generated, showing all tweets near Alberta assigned to the damage- or the donation-topic colored with the sentiment (Fig. 4.2). In the area very close to the affected area, the sentiment mainly seems to be negative in both topics. In the surrounding areas, there are more positive tweets contained in the donation topic. Next, the density based clusters for the tweets related to damage and the tweets related to donation are generated and analyzed to uncover topic hot spots. As we infer the user's location if the tweet location is not available, this clusters are influenced by the fact that all user tweets from a city are mapped to the same spot retrieved by the OpenStreetMap API. Nonetheless, this clustering can be an indicator for topic hot spots. By examining the hot spots for damage-related topics in the affected area, we could find useful information about the situation on side regarding power supply, bridge closures and pictures of flooded areas.

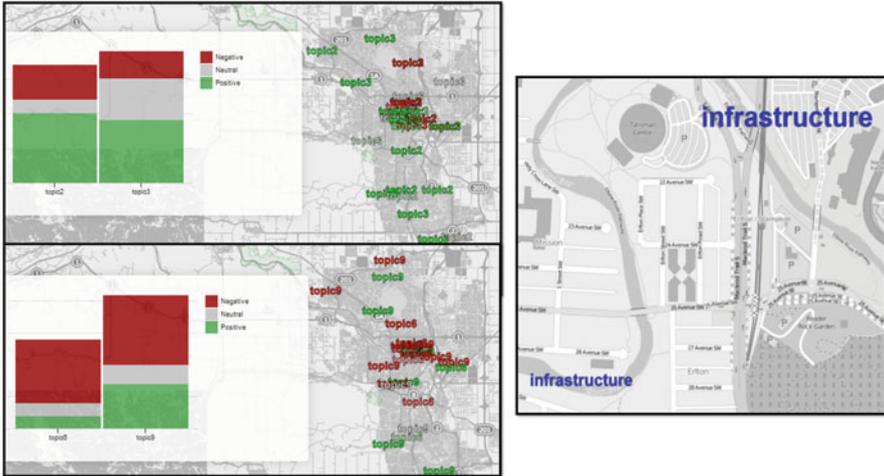


Fig. 4.2 Left: Sentiment map for tweets assigned to the damage-topics (bottom) and donation-topics (top). Right: Identified damage-cluster near the elbow river

Conclusions

In this paper, we propose a robust framework that monitors and analyzes streams from social media and is adaptable by the end-user without the need for complex tuning. The system is based on state-of-the-art text mining technologies, including a convolutional neural network for sentiment detection and an extension of the Latent Dirichlet Allocation functionality to identify latent topics from tweets. Identified topics or relevant keywords are visualized on a map together with their respective sentiment. We processed the text and enriched it with several meta features, such as the location to be able to extract different information types during the analysis and set the analysis focus to different aspects.

Acknowledgements The first author has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No 730082. The second author has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No 653811.

References

- Ashktorab, Z., Brown, C., Nandi, M., & Culotta, A. (2014). Tweedr: Mining twitter to inform disaster response. In *Iscram*, Toulouse.
- Blei, D. M., Ng, A. Y., Jordan, M. I., & Lafferty, J. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3, 993–1022.

- Caragea, C., Silvescu, A., & Tapia, A. H. (2016). Identifying informative messages in disaster events using convolutional neural networks. In *International Conference on Information Systems for Crisis Response and Management*.
- Imran, M., & Castillo, C. (2015). Towards a data-driven approach to identify crisis-related topics in social media streams. In *Proceeding of the 24th International Conference on World Wide Web* (pp. 1205–1210). New York: ACM. Retrieved from <http://doi.org/10.1145/2740908.2741729>
- Imran, M., Castillo, C., Lucas, J., Meier, P., & Vieweg, S. (2014). AIDR: Artificial intelligence for disaster response. In *Proceedings of the 23rd International Conference on World Wide Web* (pp. 159–162). New York: ACM. Retrieved from <http://doi.org/10.1145/2567948.2577034>
- Imran, M., Castillo, C., Diaz, F., & Vieweg, S. (2015). Processing social media messages in mass emergency: A survey. *ACM Computing Surveys* 47(4). <https://doi.org/10.1145/2771588>
- Kireyev, K., Palen, L., & Anderson, K. (2009). Applications of topics models to analysis of disaster-related twitter data, 11 Dec 2009.
- Nguyen, D. T., Joty, S. R., Imran, M., Sajjad, H., & Mitra, P. (2016). Applications of online deep learning for crisis response using social media information. CoRR, abs/1610.01030. Retrieved from <http://arxiv.org/abs/1610.01030>
- Olteanu, A., Castillo, C., Diaz, F., & Vieweg, S. (2014). Crisislex: A lexicon for collecting and filtering microblogged communications in crises. In *International AAI Conference on Web and Social Media*. Retrieved from <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8091>
- Olteanu, A., Vieweg, S., & Castillo, C. (2015). What to expect when the unexpected happens: Social media communications across crises. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 994–1009). New York: ACM. Retrieved from <http://doi.acm.org/10.1145/2675133.2675242>
- Wang, S., Chen, Z., Fei, G., Liu, B., & Emery, S. (2016). Targeted topic modeling for focused analysis. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining – KDD’16* (pp. 1235–1244). Retrieved from <http://dl.acm.org/citation.cfm?doid=2939672.2939743>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

