Crisis Management Knowledge from Social Media

Karl Kreiner
AIT Austrian Institute of Technology GmbH, Safety & Security Department
Graz, Austria
karl.kreiner@ait.ac.at

Aapo Immonen
The University of Eastern Finland
Department of Public Health
Kuopio, Finland
Aapo.Immonen@fcg.fi

Hanna Suominen
NICTA, The Australian National University, and University of Canberra
Canberra, ACT, Australia
hanna.suominen@nicta.com.au

ABSTRACT

More and more crisis managers, crisis communicators and laypeople use Twitter and other social media to provide or seek crisis information. In this paper, we focus on retrospective conversion of human-safety related data to crisis management knowledge. First, we study how Twitter data can be classified into the seven categories of the United Nations Development Program Security Model (i.e., Food, Health, Politics, Economic, Personal, Community, and Environment). We conclude that these topic categories are applicable, and supplementing them with classification of individual authors into more generic sources of data (i.e., Official authorities, Media, and Laypeople) allows curating data and assessing crisis maturity. Second, we introduce automated classifiers, based on supervised learning and decision rules, for both tasks and evaluate their correctness. This evaluation uses two datasets collected during the crises of Queensland floods and NZ Earthquake in 2011. The topic classifier performs well in the major categories (i.e., 120–190 training instances) of Economic (F = 0.76) and Community (F = 0.67) while in the minor categories (i.e., 0–60 training instances) the results are more modest ($F \leq 0.41$). The source classifier shows excellent results ($F \geq 0.83$) in all categories.

Categories and Subject Descriptors

H.3.2. [**Information Storage and Retrieval**] Information Search and Retrieval – *search process*

H.4.1 [**Information Systems and Applications**] Office Automation – *workflow management*

General Terms

Design, Experimentation

Keywords

Data processing, Information retrieval, Knowledge management, Social network services.

ADCS '13, December 05 – 06 2013, Brisbane, QLD, Australia

http://dx.doi.org/10.1145/2537734.2537740

1. INTRODUCTION

Crisis management responds to an emergency or disaster with the purpose of saving lives, property, and environment [1]. It uses *social media*, both as a source of data for situational awareness, cognitive empowerment, and decision support and as a distribution channel of official authorities, media, and laypeople [2]. This allows people's voice to be heard as interactive mobile sensors who post new messages on potential early indicators, and peer review the previously posted messages [3, 4]. Social media can also provide wider and deeper information, for example, by better representing rural areas and particular locations than mass media [3]. In California in Oct 2008, when wildfires destroyed 500,000 acres of land and 1,500 homes, 76% (36%) of people sought (provided) information on the Internet [3]. In Kenya after the post-election fallout in 2008, 45,000 people used a maps application on the Internet to seek and provide information on incidents of violence and peace efforts [5]. In Haiti in Jan 2010, when an earthquake killed 230,000 people, volunteers gathered input for this maps application from the Internet and authorities then used its output to direct assistance [6]. In Feb 2010, more than 4.5 million social media messages either described or curated information related to a Chilean earthquake, which killed 723 people and damaged 370,000 homes [4]. During the London riots in Aug 2011 in the UK, the use of a messenger system played a key role both in the organisation of rioting and riot cleaning [7].

Search engines are needed to leverage collective wisdom in social media by converting the gamut of data into knowledge to crisis management knowledge [8, 9]. For example, *topic modeling* on social media has been studied generally [10] and specifically for disaster-related data [9]. The results show that tailored topic models are able to distinguish distinct disasters of similar nature in nearby locations (Sumatran earthquake versus Samoan tsunami) and support exploring prominent topics. By considering time, topic modeling has been extended to cover the *detection of emerging topics* on social media [11]. Applications of these methods to *event detection* in real time include developing and evaluating an earthquake detector [12] and a system for measuring evolving events and describing this evolution with the topic of influenza A H1N1 [13]. Extensions address *sentiment analysis* to determine authors' opinions [14]. Finally, potential contributions of social media equipped with intelligent search engines to official authorities, media, and laypeople's situational awareness has been analysed in the cases of the Oklahoma grass fires and Red River floods in the USA in Mar–Apr 2009 [15]. The envisioned search capabilities consist of identifying locations; generating general warnings and those customised to a given type of natural disaster; planning responses to warnings, including evacuation, sheltering, and animal management; analysing other environmental conditions; providing advice; and reporting damages and injuries.

This paper studies how the *United Nations Development Program Security Model* (UNDPSM) [16] – used by most crisis managers and applicable to a wide range of (if not all) crises – is applicable to (automated) retrospective analyses of social media.

2. MATERIALS AND METHODS

Two datasets originated from *Twitter* were used (Table 1). Twitter is a real-time messaging network established in 2006. In Jun 2011, Twitter's more than 300 million registered users (150 million active users) sent 200 million messages per day (vs. 65 million in Jun 2010 and two million in Jan 2009) which is equivalent to ten million pages or over 31 years of reading time [17]. Each of these *Tweets* was up to 140 characters long and could have metadata for additional information, deeper context, and embedded media. Some Tweets referred to Twitter users and topics with symbols @ (e.g., *@QldFire* is the official Twitter account of Queensland Fire and Rescue Service) and # (e.g., #floods), respectively. The first dataset (i.e., *AU_floods*) was collected during major flooding in early 2011 in Queensland, AU. The flood's area was 1,000,000 km^2, which is more than the combined area of Texas and New Mexico in the USA, or France and Germany in Europe. The second dataset (i.e., *NZ_earthquake*) was collected in Feb 2011, when a major earthquake hit the city of Christchurch in NZ leaving 185 people dead, and the estimated costs of re-construction of USD 17–25 billion [18]. Data examples include *Any1 with boat in Chelmer nr rosebery tce? dad has been stuck for over 4hrs & we can't get in touch #bneflood #qldfloods* and *the scene outside riparian plaza now! massive flood water encroachment #qldfloods http://twitpic.com/3p9lxd.*

We followed the *annotation process* [19] with the theory for annotation being initialised by assigning a topic to each Tweet (i.e., *topic classification*) on the basis of the UNDPSM definitions of seven areas of threats (i.e., *Food*, *Health*, *Politics*, *Economic*, *Personal*, *Community*, and *Environment*) [16] and the additional category of *Unknown* for any other content. We used two annotators trained and managed by one crisis management expert. The annotators worked independently of each other and used *inductive content analysis* [20] to assign a topic for each Tweet. For each topic assignment, they gave an certainty score on a five-point scale of *certain–relatively certain–neutral–relatively uncertain–uncertain* Our annotation interface was a *Microsoft Excel* spreadsheet with the first column for Tweets, columns two through nine for the seven topics, the tenth column for certainty scores, and the eleventh column for free-text comments. We measured inter-annotator agreement via the *R 2.6.2* implementation of *Cohen's κ* with values < 0.0, 0.0–0.2, 0.2–0.4, 0.4–0.6, 0.6–0.8, and 0.0–1.0 for poor, slight, fair, moderate, substantial, and almost perfect agreement, respectively [21].

We iterated until the agreement was almost perfect (i.e., six iterations). For the first four iterations (Table 2), we randomly selected 340 Tweets from the AU_floods dataset that remained after relevance filtering. For relevance filtering, we first excluded forwarded Tweets and then, based on regular expressions, included Tweets containing a question [i.e., sentences that begin with a question word (e.g., *where*, *when*, *why*, *does*)], instructions (i.e., sentences that begin with a verb or end with an exclamation mark), reference to a location, or links to popular image services. In the inter-annotator comparison for the first iteration, we observed confusion between *Community*, *Environment*, *Personal*, and *Unknown*. After clarifying the guideline by annotators' discussion, three iterations were conducted but the confusion between the categories remained. The fourth guideline was: *Food*: any reference to solid or liquid food (containing water) no matter if it affects an individual, a group of persons or a community as a whole. *Health*: any reference to health-related issues no matter if it affects an individual, a group of persons or a community as a whole. *Politics*: any reference to content that directly affects the political security of a person. *Economy*: any reference to monetary units, or topics that directly influence economic issues (jobs – no volunteer work, donations). *Personal*: any reference to man-made private property of any kind (including animals); an individual (the author itself is a reference too) or a specific group of people; help is required by individuals, offered to individuals, or offered to a group of specific people. When in doubt, if the property is private or public, assume the property to be private. *Community*: any reference to man-made public property (cities, communities are public property as well or property that is accessible by the public) of any kind, when in doubt consider property being private; if the people of a community are addressed as a whole; if help in any way is offered to a community or a non-specific group of people. *Environment*: any reference to non-man-made structures (e.g., hills, lakes, and oceans), natural phenomena (e.g., rain and storm) or natural effects caused by man (e.g., hazardous material and radiation) that directly impose a security risk to individuals, a group of persons or a community as a whole. *Unknown*: if a link is provided within the Tweet try to apply the seven categories to the content of the first page but if this fails, mark as Unknown. To improve agreement, we clarified and simplified the category definitions and cascaded the process:

> If the message is relevant to food quality or supplies, assign it to the category of *Food*. Otherwise if it refers to matters of private or public health, assign it to the category of *Health*. Otherwise if it contains political commentary, assign it to the category of *Politics*. Otherwise if it offers or reports help or advice, assign it to the category of *Community*. Otherwise if it describes people seeking help, assign it to the category of *Personal*. Otherwise if it calls for donations, assign it to the category of *Economic*. Otherwise if it includes information on natural hazards, assign it to the category of *Environment*. Otherwise assign it to *Unknown*.

We used this final guideline during the last two iterations. We selected randomly new data (500 AU_floods Tweets and 500 NZ_earthquake Tweets). This resulted in almost perfect inter-annotator agreement: Cohen's κ was 0.72 after the fifth iteration, and after cleaning obvious errors as the sixth iteration, Cohen's κ was 0.81. We formed the final annotation from the annotators' discussion for agreement. *Community* and *Economic* were clearly the most common categories in both datasets (Fig. 1).

We supplemented this classification with *source classification* of individual authors into three more generic categories in order to allow curating data and assessing crisis maturity:

> *Media*: Newspapers; online magazines; television and radio stations; and other media assets. *Official authorities and non-profit organisations*: Emergency agencies; ministries; police; fire services; military sources; non-profit organisations (e.g., Red Cross or Green Cross); and volunteer organisations. *Laypeople*: Anything that does not fit into category media or official authorities including private persons and/or businesses.

As data for this classification, we used the author's website. We collected a balanced sample of 600 websites, that is, 200 sites per category, manually. First, we performed a Google search for media outlets in Great Britain, Canada, South Africa, AU, NZ, and USA. Second, we performed a Google search for emergency agencies in English-speaking countries. Third, we added military, police, fire and rescue services, as well as official sources like ministries. Finally, we chose 200 links from the NZ dataset including links to private websites, blogs, and Facebook profiles.

We built automated *topic and source classifiers*. The topic classifier converted each message into a feature vector via linguistic processing, including *shorthand expansion*; *lemmatisation and part-of-speech tagging*; *replacement of hyponyms with their hypernyms*; and *sentiment analysis*. Lemmatisation and part-of-speech tagging used a tailored Twitter Part-of-speech Tagger. We used the *WordNet Lexical Database for English module* of *NLTK* to replace of hyponyms (i.e., more specific words) with their hypernyms (i.e., more general words). This was used to recognise references to food, currencies, geographic locations, natural phenomena, and possession. We supplemented the tag set with references to images (e.g., *http://yfrog.com/imagename*), numbers (e.g., *120.00*), money (e.g., *$*), geographic locations (e.g., *Teneriffe*), Twitter users (e.g., *@username*), Twitter hash tags (e.g., *#thebigwet*), uniform resource locators (e.g., *http://twitter.com/*), and shorthand (e.g., *RSPCA*). We used the *Regressive Imagery Dictionary on NodeBox* to extract psychological features from the text. Each feature had a binary value, indicating its presence or absence. First, all features were used. Second, this was narrowed down to the top 72 features. The source classifier was built similarly. We wrote a parser that automatically loaded the link to the user's website and investigated the links of the loaded page. Based on 598 websites collected earlier, we extracted the top terms found for each category (i.e., typical keywords included *weather*, *business*, *news*, *latest*, *breaking*, *sports*, *magazine*, *opinion*, and *weekly* for *Media*; *emergency*, *agency*, *fire*, *police*, *government*, *rescue*, *assistance*, and *federal* for *Official authorities and non-profit organisation*; and *blog*, *I*, *me*, *am*, *blogroll*, *welcome*, *profile*, *product*, *products*, and *commercial* for *Laypeople*). We inspected the link text, since often link name gives hints on the website content (e.g., the domain *gov* indicates, that the website is run by a government agency). We derived all binary features and narrowed down to the top 123 features. Experiments were conducted using the *Naïve Bayes* classifier of the *Orange Machine Learning Toolkit V2.6*. For performance evaluation, we used the *number of true positives* (*TPc*), *false positives* (*FPc*), *true negatives* (*TNc*), *false negatives* (*FNc*), *precision* (*Pc*), *recall* (*Rc*), and F1 (*Fc*) for each category *c* with *leave-one-out cross-validation* to define the measure values. We compared the results to always assigning the majority class.

3. RESULTS AND DISCUSSION

The source classifier shows excellent results ($F \geq 0.83$) in all categories and the topic classifier performs well in the major categories (i.e., 120–190 training instances) of *Economic* ($F = 0.76$) and *Community* ($F = 0.67$) while more modestly ($F \leq 0.41$) in the minor categories (i.e., 0–60 training instances). Because our datasets did not include any messages annotated to *Politics*, we were not able to automatically classify messages to this category.

Search engines built based on the datasets and methods of this paper could serve educational purposes in crisis management and communication after a crisis. For example, the distribution of messages across the topics could demonstrate the nature of the crisis – does the crisis relate more strongly to the economic situation or environmental stability? When applied periodically over time to produce cross-sectional statuses, then combined to longitudinal trends, these engines could reveal early warnings for certain developments during a crisis. A crisis can be described as a sigma curve of its phases on the *x*-axis and the extent of crisis management activities on the *y*-axis [1]. The curve begins from the faint emergence of early indicators to the normality (e.g., people's experiences or breaking news). This is continued by the early warnings, notification, and emergency or disaster management, which co-occurs with the peak of the curve. Finally, the curve fades out back to the normality. Search engines hold the potential for tracking the sigma curve by retrospective analysis of social media. Finally, the retrospective analysis capability could support studying effects of crisis communication.

Table 1. Datasets[a]

	AU_floods	NZ_earthquake
Crisis	Floods	Earthquake
Location	Queensland, AU	Christchurch, NZ
Time frame	Jan–Mar 2011	Feb 2011
Source	TwapperKeeper, a public archive at twapperkeeper.com	Twitter
Gatherer	Jean Burgess at twitter.com/jeanburgess	Us using Twitter Application Programmer's Interface
Search	#qldfloods, #qldflood	#eqnz, #chch, #earthquake, 25 km from Christchurch
Data	Message, date, user name	Message, date, full user profile
Tweets	48,016	52,602
Authors	17,515	9,419[b]
Words[c] (unique)	831,947 (64,742)	856,586 (75,849)

a. Please email the authors if you wish to access the datasets.
b. 10,717 user profiles, however some profiles were inaccessible (e.g., deleted spam account) in the course of the study
c. via Apple Terminal commands without pre-processing of text

Table 2. The first four iterations of the annotation process with 340 tweets from the AU_flood dataset

Iteration	1	2	3	4
Cohen's κ	0.32	0.30	0.54	0.49
DA No. (%)				
Com–Env	34 (10%)	58 (17%)	13 (3.8%)	16 (4.7%)
Com–Per	27 (7.9%)	25 (7.4%)	25 (7.4%)	29 (8.5%)
Com–UK	21 (6.2%)	18 (5.3%)	28 (8.2%)	9 (2.6 %)

% refers to the comparison with the 340 annotation decisions for each annotator; DA. = disagreements, Com = community, Env = environment, Per = personal, UK = unknown

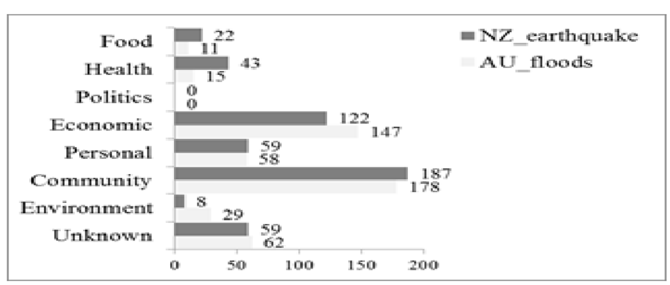

Figure 1. Topic distribution of the 500 + 500 Tweets

Table 3. Performance of our method (baseline) in topic classification (top) and source classification (bottom)

Category	*TPc %*	*FPc %*	*TNc %*	*FNc %*	*Pc*	*Rc*	*Fc*
Unknown	0.31 (0.00)	0.08 (0.00)	0.92 (0.00)	0.69 (1.00)	0.34 (0.00)	0.24 (0.00)	0.28 (0.00)
Environment	0.30 (0.00)	0.05 (0.00)	0.95 (1.00)	0.70 (1.00)	0.20 (0.00)	0.29 (0.00)	0.24 (0.00)
Community	0.74 (1.00)	0.25 (1.00)	0.75 (0.00)	0.26 (0.00)	0.62 (0.36)	0.73 (1.00)	0.67 (0.53)
Personal	0.36 (0.00)	0.07 (0.00)	0.93 (1.00)	0.64 (1.00)	0.42 (0.00)	0.35 (0.00)	0.39 (0.00)
Economic	0.78 (0.00)	0.10 (0.00)	0.90 (1.00)	0.22 (1.00)	0.74 (0.00)	0.78 (0.00)	0.76 (0.00)
Health	0.31 (0.00)	0.02 (0.00)	0.98 (1.00)	0.69 (1.00)	0.62 (0.00)	0.31 (0.00)	0.41 (0.00)
Food	0.18 (0.00)	0.01 (0.00)	0.99 (1.00)	0.82 (1.00)	0.50 (0.00)	0.18 (0.00)	0.27 (0.00)
Class-size weighted average	0.59 (0.36)	0.13 (0.36)	0.85 (0.51)	0.40 (0.62)	0.58 (0.13)	0.58 (0.36)	0.57 (0.19)
Media	0.81 (0.00)	0.04 (0.00)	0.96 (1.00)	0.29 (1.00)	0.91 (0.00)	0.81 (0.00)	0.86 (0.00)
Official authorities…	0.88 (0.00)	0.03 (0.00)	0.97 (1.00)	0.12 (1.00)	0.95 (0.00)	0.88 (0.00)	0.92 (0.00)
Laypeople	0.90 (1.00)	0.05 (1.00)	0.85 (0.00)	0.15 (0.00)	0.77 (0.34)	0.90 (1.00)	0.83 (0.51)
Class-size weighted average	0.86 (0.34)	0.04 (0.34)	0.93 (0.67)	0.19 (0.67)	0.88 (0.11)	0.86 (0.34)	0.87 (0.17)

Acknowledgements
NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program. KK and AI were visiting PhD students at NICTA and HS was their supervisor.

4. REFERENCES

[1] Immonen, A., Bali, R. K., and Naguib, R. 2009. Towards a knowledge-based conceptual model for post-crisis public health scenarios. In *Proc. HNICEM 4th International Conference Humanoid, Nanotechnology, Information Technology Communication and Control, Environment and Management* (The Institute of Electrical and Electronics Engineers Inc. (IEEE), Manila, Philippines).

[2] Palen L. 2008. Online social media in crisis events. *EDUCAUSE Quarterly* 31, 3 (Jul–Sep 2008).

[3] Sutton, J., Palen, L., and Shklovski, I. 2008. Backchannels on the front lines: emergent use of social media in the 2007 Southern California wildfires. *In Proc. ISCRAM 2008 Information Systems for Crisis Response and Management Workshop* (Washington DC, USA, 624–632).

[4] Mendoza, M., Poblete, B., and Castillo, C. 2010. Twitter under crisis: can we trust what we RT? In *SOMA'10 Proc. 1st Workshop on Social Media Analytics* (ACM, Washington, DC, USA).

[5] Ushahidi. 2012. *About Us*. goo.gl/mA4aJA

[6] Heinzelman, J., and Waters, C. 2010. Crowdsourcing crisis information in disaster-affected Haiti. *United States Institute of Peace Special Report* (Washington DC, USA).

[7] Yelaja, P. 2011. U.K. riots reveal social media double standard. *CBC News*. goo.gl/kf43X

[8] Allan, J., Aslam, J., Belkin, N., Buckley, C., Callan, J., Croft, B., et al. 2002. Challenges in information retrieval and language modeling: report of a workshop held at the Center for Intelligent Information Retrieval, University of Massachusetts Amherst, September 2002. *SIGIR Forum* 37, 1, 31-47.

[9] Kireyev, K., Palen, L., and Anderson,K. M. 2009. Applications of topics models to analysis of disaster-related Twitter data. In *NIPS Workshop on Applications for Topic Models: Text and Beyond* (Amherst, MA, USA).

[10] Hong, L., and Davison, B. D. 2010. Empirical study of topic modeling in Twitter. In *Proc. SOMA'10 1st Workshop on Social Media Analytics* (ACM, Washington, DC, USA).

[11] Cataldi, M., Di Caro, L., and Schifanella, C. 2010. Emerging topic detection on Twitter based on temporal and social terms evaluation. In *Proc. MDMKDD'10 Tenth International Workshop on Multimedia Data Mining* (ACM, New York, NY, USA, article no. 4).

[12] Sakaki, T., Okazaki, M., and Matsuo, Y. 2010. Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proc. WWW '10 19th International Conference on World Wide Web* (ACM, New York, NY, USA, 851–860).

[13] Signorini, A., Segre, A. M., and Polgreen, P. M. 2011. Public concern in the U.S. during the influenza A H1N1 pandemic. *PLoS ONE* 6, 5 (May 2011), e19467.

[14] Pak, A., and Paroubek, P. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *Proc. LREC'10 Seventh International Conference on Language Resources and Evaluation* (European Language Resources Association (ELRA),Valletta, Malta).

[15] Vieweg, S., Hughes, A. L., Starbird, K.., and Palen,L. 2010. Microblogging during two natural hazards events: what Twitter may contribute to situational awareness. *In Proc. CHI 2010: Crisis Informatics* (ACM, Atlanta, GA, USA, 1079–1088).

[16] The United Nation. 1994. *Human Development Report*. goo.gl/1bdY4V

[17] Twitter Blog. 2011. *#numbers* and *200 Million Tweets per Day*. goo.gl/Z9wqk and goo.gl/ruVuZ

[18] Buhayar, N., Greber, J., and Saminather, N. 2011. New Zealand's earthquake may become costliest insured disaster since 2008. *Bloomberg*. goo.gl/z3L9e

[19] Hovy, E. 2010. *Creating Training Material for Health Informatics: Toward a Science of Annotation*. goo.gl/o1nZZX

[20] Hsieh, H.-F., and Shannon, S. E.. 2005. Three approaches to qualitative content analysis. *Qual Health Res.* 15, 9 (Nov 2009), 1277–1288.

[21] Landis, J. R., and Koch, G. G. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 1 (Mar 1977), 159–174..