

A Cross-Domain Exploratory Analysis of Social Media Data During Crisis Events

Hernan Sarmiento

Department of Computer Science
University of Chile
Santiago, Chile
hsarmien@dcc.uchile.cl

Barbara Poblete

Department of Computer Science
University of Chile
Santiago, Chile
bpoblete@dcc.uchile.cl

ABSTRACT

CCS CONCEPTS

• **Information systems** → **Data stream mining**; *Spatial-temporal systems*;

KEYWORDS

Emergency Situations; Citizen Sensors; Social Media

ACM Reference Format:

Hernan Sarmiento and Barbara Poblete. 2019. A Cross-Domain Exploratory Analysis of Social Media Data During Crisis Events. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

2 DATA ACQUISITION

- CrisisLex and CrisisNlp.
- XX dataset with YY instances and ZZ events.

3 DATA PREPARATION

- Remove duplicates
- Remove data without metadata

3.1 Detecting Languages

3.2 Adding country information

3.2.1 Adding primary languages.

3.3 Merging Labels

3.4 Adding Crisis Categories

4 EXPLORATORY ANALYSIS

4.1 Computing Features

4.1.1 Content Features.

4.1.2 Grammatical Features.

4.1.3 Entities-Based Features.

4.1.4 Emotional Features.

4.2 Hazard Categories Analysis

4.3 Hazard Sub Categories Analysis

4.4 Hazard Types Analysis

- 82% english data and 8% spanish and italian each.
- Most events occurred in United States

ACKNOWLEDGMENTS

This work has been partially funded by the Millennium Institute for Foundational Research on Data.

REFERENCES

- [1] Fabian Abel, Claudia Hauff, Geert-Jan Houben, Richard Stronkman, and Ke Tao. 2012. Twitcident: fighting fire with information from social web streams. In *Proceedings of the 21st International Conference on World Wide Web*. ACM, 305–308.
- [2] Hila Becker, Mor Naaman, and Luis Gravano. 2011. Beyond Trending Topics: Real-World Event Identification on Twitter. *ICWSM 11, 2011* (2011), 438–441.
- [3] Axel Bruns and Jean E Burgess. 2012. Local and global responses to disaster: #eqnz and the Christchurch earthquake. In *Disaster and emergency management conference, conference proceedings*, Vol. 2012. AST Management Pty Ltd, 86–103.
- [4] Mark A Cameron, Robert Power, Bella Robinson, and Jie Yin. 2012. Emergency situation awareness from twitter for crisis management. In *Proceedings of the 21st International Conference on World Wide Web*. ACM, 695–698.
- [5] Cornelia Caragea, Nathan McNeese, Anuj Jaiswal, Greg Traylor, Hyun-Woo Kim, Prasenjit Mitra, Dinghao Wu, Andrea H Tapia, Lee Giles, Bernard J Jansen, et al. 2011. Classifying text messages for the Haiti earthquake. In *Proceedings of the 8th international conference on information systems for crisis response and management (ISCRAM2011)*. Citeseer.
- [6] Lowell Juilliard Carr. 1932. Disaster and the sequence-pattern concept of social change. *Amer. J. Sociology* 38, 2 (1932), 207–218.
- [7] Carlos Castillo. 2016. *Big crisis data: social media in disasters and time-critical situations*. Cambridge University Press.
- [8] UN DHA. 1992. Internationally agreed glossary of basic terms related to disaster management. *UN DHA (United Nations Department of Humanitarian Affairs), Geneva* (1992).
- [9] Mark Graham, Scott A Hale, and Devin Gaffney. 2014. Where in the world are you? Geolocation and language identification in Twitter. *The Professional Geographer* 66, 4 (2014), 568–578.
- [10] Jheser Guzman and Barbara Poblete. 2013. On-line relevant anomaly detection in the Twitter stream: an efficient bursty keyword detection model. In *Proceedings of the acm sigkdd workshop on outlier detection and description*. ACM, 31–39.
- [11] Brent Hecht, Lichan Hong, Bongwon Suh, and Ed H Chi. 2011. Tweets from Justin Bieber's heart: the dynamics of the location field in user profiles. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 237–246.
- [12] Quynh Huang, Guido Cervone, Duangyang Jing, and Chaoyi Chang. 2015. DisasterMapper: A CyberGIS framework for disaster management using social media data. In *Proceedings of the 4th International ACM SIGSPATIAL Workshop on Analytics for Big Geospatial Data*. ACM, 1–6.
- [13] Muhammad Imran, Carlos Castillo, Fernando Diaz, and Sarah Vieweg. 2015. Processing social media messages in mass emergency: A survey. *ACM Computing Surveys (CSUR)* 47, 4 (2015), 67.
- [14] Muhammad Imran, Carlos Castillo, Ji Lucas, Patrick Meier, and Sarah Vieweg. 2014. Aidr: Artificial intelligence for disaster response. In *Proceedings of the companion publication of the 23rd international conference on World wide web*

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
Conference'17, July 2017, Washington, DC, USA
© 2019 Copyright held by the owner/author(s).
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

- companion. International World Wide Web Conferences Steering Committee, 159–162.
- [15] Muhammad Imran, Shady Elbassuoni, Carlos Castillo, Fernando Diaz, and Patrick Meier. 2013. Extracting information nuggets from disaster-Related messages in social media.. In *ISCRAM*.
 - [16] Janani Kalyanam, Mauricio Quezada, Barbara Poblete, and Gert Lanckriet. 2016. Prediction and Characterization of High-Activity Events in Social Media Triggered by Real-World News. *PloS one* 11, 12 (2016), e0166694.
 - [17] Sarvnaz Karimi, Jie Yin, and Cecile Paris. 2013. Classifying microblogs for disasters. In *Proceedings of the 18th Australasian Document Computing Symposium*. ACM, 26–33.
 - [18] Yury Kryvasheyev, Haohui Chen, Nick Obradovich, Esteban Moro, Pascal Van Hentenryck, James Fowler, and Manuel Cebrian. 2016. Rapid assessment of disaster damage using social media activity. *Science advances* 2, 3 (2016), e1500779.
 - [19] Shamanth Kumar, Geoffrey Barbier, Mohammad Ali Abbasi, and Huan Liu. 2011. TweetTracker: An Analysis Tool for Humanitarian and Disaster Relief.. In *ICWSM*.
 - [20] Rui Li, Kin Hou Lei, Ravi Khadiwala, and Kevin Chen-Chuan Chang. 2012. Tedas: A twitter-based event detection and analysis system. In *Data engineering (icde), 2012 IEEE 28th international conference on*. IEEE, 1273–1276.
 - [21] John Lingad, Sarvnaz Karimi, and Jie Yin. 2013. Location extraction from disaster-related microblogs. In *Proceedings of the 22nd international conference on world wide web*. ACM, 1017–1020.
 - [22] Nicola Lunardon, Giovanna Menardi, and Nicola Torelli. 2014. ROSE: A Package for Binary Imbalanced Learning. *R Journal* 6, 1 (2014).
 - [23] Jazmine Maldonado, Jheser Guzman, and Barbara Poblete. 2017. A Lightweight and Real-Time Worldwide Earthquake Detection and Monitoring System Based on Citizen Sensors. In *Proceedings of the Fifth Conference of Human Computation and Crowdsourcing*. AAAI, 137–146.
 - [24] Kanti V Mardia. 1970. Measures of multivariate skewness and kurtosis with applications. *Biometrika* 57, 3 (1970), 519–530.
 - [25] Marcelo Mendoza, Barbara Poblete, and Carlos Castillo. 2010. Twitter Under Crisis: Can we trust what we RT?. In *Proceedings of the first workshop on social media analytics*. ACM, 71–79.
 - [26] Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Sarah Vieweg. 2014. CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises.. In *ICWSM*.
 - [27] Alexandra Olteanu, Sarah Vieweg, and Carlos Castillo. 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*. ACM, 994–1009.
 - [28] Liza Potts, Joyce Seitzinger, Dave Jones, and Angela Harrison. 2011. Tweeting disaster: hashtag constructions and collisions. In *Proceedings of the 29th ACM international conference on Design of communication*. ACM, 235–240.
 - [29] Christian Reuter and Marc-André Kaufhold. [n. d.]. Fifteen years of social media in emergencies: a retrospective review and future directions for crisis informatics. *Journal of Contingencies and Crisis Management* ([n. d.]).
 - [30] Kevin Stowe, Michael J Paul, Martha Palmer, Leysia Palen, and Kenneth Anderson. 2016. Identifying and categorizing disaster-related tweets. In *Proceedings of The Fourth International Workshop on Natural Language Processing for Social Media*. 1–6.
 - [31] Sayan Unankard, Xue Li, and Mohamed A Sharaf. 2015. Emerging event detection in social networks with location sensitivity. *World Wide Web* 18, 5 (2015), 1393–1417.
 - [32] Maximilian Walther and Michael Kaissner. 2013. Geo-spatial Event Detection in the Twitter Stream.. In *ECIR*. Springer, 356–367.
 - [33] Kazufumi Watanabe, Masanao Ochi, Makoto Okabe, and Rikio Onai. 2011. Jasmine: a real-time local-event detection system based on geolocation information propagated to microblogs. In *Proceedings of the 20th ACM international conference on Information and knowledge management*. ACM, 2541–2544.
 - [34] Jie Yin, Sarvnaz Karimi, and John Lingad. 2014. Pinpointing locational focus in microblogs. In *Proceedings of the 2014 Australasian Document Computing Symposium*. ACM, 66.
 - [35] Jie Yin, Sarvnaz Karimi, Bella Robinson, and Mark Cameron. 2012. ESA: emergency situation awareness via microbloggers. In *Proceedings of the 21st ACM international conference on Information and knowledge management*. ACM, 2701–2703.