Mining Social Networks to Learn about Rumors, Hate Speech, Bias and Polarization - Abstract

Bárbara Poblete[a]

[a]*University of Chile, Chile*

Abstract

Online social networks are a rich resource of unedited user-generated multimedia content. Buried within their day-to-day chatter, we can find breaking news, opinions and valuable insight into human behaviour, including the articulation of emerging social movements. Nevertheless, in recent years social platforms have become fertile ground for diverse information disorders and hate speech expressions. This situation poses an important challenge to the extraction of useful and trustworthy information from social media.

In this talk I provide an overview of existing work in the area of social media information credibility, starting with our research in 2011 on rumor propagation during the massive earthquake in Chile in 2010 [1]. I discuss, as well, the complex problem of automatic hate speech detection in online social networks. In particular, how our review of the existing literature in the area shows important experimental errors and dataset biases that produce an overestimation of current state-of-the-art techniques [2]. Especifically, these issues become evident at the moment of attempting to apply these models to more diverse scenarios or to transfer this knowledge to languages other than English.

As a particular way of dealing with the need to extract reliable information from online social media, I talk about two applications, Twically [3] and Galean [4]. These applications harvest collective signals created from social media text to provide a broad view of natural disasters and real-world news, respectively.

Keywords

Online social networks, information credibility, hate speech

Biographical Sketch

Dr. Barbara Poblete is an Associate Professor at the Computer Science (CS) Dept. of the University of Chile. She is also a Researcher at the Millennium Institute for Foundational Research on Data, where she co-leads the ”Fake News and Misinformation” multidisciplinary research group. Formerly, she was a Researcher at Yahoo! Labs. Her current research areas include Applied Machine Learning, Data Mining, Experimental Reproducibility, Online Social Networks Analysis, Hate Speech Detection, Crisis Informatics and Information Retrieval. Her series of work on ”information credibility in social media” (starting in 2010) have been widely cited and were the first scientific studies addressing online misinformation in social networks. Her research on this and other topics has appeared in Scientific American Magazine, The Wall Street Journal, Slate Magazine, The Huffington Post, BBC News and NPR, among others.

OHARS'20: Workshop on Online Misinformation- and Harm-Aware Recommender Systems, September 25, 2020, Virtual Event

References

[1] C. Castillo, M. Mendoza, B. Poblete, Information credibility on twitter, in: Proceedings of the 20th International Conference on World Wide Web, WWW '11, Association for Computing Machinery, New York, NY, USA, 2011, p. 675–684. doi:10.1145/1963405.1963500.

[2] A. Arango, J. Pérez, B. Poblete, Hate speech detection is not as easy as you may think: A closer look at model validation (extended version), Information Systems (2020) 101584. doi:10.1016/j.is.2020.101584.

[3] B. Poblete, J. Guzmán, J. Maldonado, F. Tobar, Robust detection of extreme events using twitter: Worldwide earthquake monitoring, IEEE Transactions on Multimedia 20 (2018) 2551–2561. doi:10.1109/TMM.2018.2855107.

[4] V. Peña-Araya, M. Quezada, B. Poblete, D. Parra, Gaining historical and international relations insights from social media: spatio-temporal real-world news analysis using twitter, EPJ Data Science 6 (2017) 25.