Toxicity, polarizations and cultural diversity in social networks.

Using machine learning and natural language processing to analyze these phenomena in social networks.

Abi Oppenheim[1], Federico Albanese[2,3], and Esteban Feuerstein[1,3]

{aoppenheim, falbanese, efeuerst}@dc.uba.ar

[1] Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina.
[2] Instituto de Cálculo, CONICET Universidad de Buenos Aires,Buenos Aires, Argentina.
[3] Instituto de Ciencias de la Computación, CONICET Universidad de Buenos Aires, Buenos Aires, Argentina.

Abstract. Social media have increased the amount of information that people consume as well as the number of interactions between them. Nevertheless, most people tend to promote their favored narratives and hence form polarized groups [1]. This encourages polarization and extremism resulting in extreme violence [2]. Against this backdrop, it is in our interest to find environments, strategies and mechanisms that allow us to reduce toxicity on social media (defining "toxicity" as a rude, disrespectful or unreasonable comment that is likely to make people leave a discussion). We address the hypothesis that a higher cultural diversity among community users reduces the toxicity of the user messages. We use Reddit as a case study, since this platform is characterized by a variety of discussion sub-forums where users debate political and cultural issues. Using community2vec [3], we generate an embedding for each community that allows us to portray users in a demographic and ideological aspect [4]. In order to analyze each user statement, we process the data with different models, thereby obtaining which are the topics of debate and what are the levels of aggressiveness and negativism in them. Finally, we will seek to corroborate the hypothesis by analyzing the relationship between the cultural diversity present in each discussion group and the toxicity found in their posts.

Keywords: Machine learning · Social media · Reddit · Data mining · Toxicity.

References

1. QUATTROCIOCCHI, Walter; SCALA, Antonio; SUNSTEIN, Cass R. Echo chambers on Facebook. Available at SSRN 2795110, 2016.

2   Abi Oppenheim

2. MILLS, Judson; ARONSON, Elliot; ROBINSON, Hal. Selectivity in exposure to information. The Journal of Abnormal and Social Psychology, 1959, vol. 59, no 2, p. 250.
3. MARTIN, Trevor. community2vec: Vector representations of online communities encode semantic relationships. En Proceedings of the Second Workshop on NLP and Computational Social Science. 2017. p. 27-31.
4. WALLER, Isaac; ANDERSON, Ashton. Quantifying social organization and political polarization in online platforms. Nature, 2021, vol. 600, no 7888, p. 264-268.