Paolo Rosso, Universitat Politècnica de València

Profiling Bots, Fake News Spreaders and Haters

Author profiling studies how language is shared by people. Stylometry techniques help in identifying aspects such as gender, age, native language, or even personality. Author profiling is a problem of growing importance, not only in marketing and forensics, but also in cybersecurity. The aim is not only to identify users whose messages are potential threats from a terrorism viewpoint but also those whose messages are a threat from a social excusion perspective because containing hate speech, cyberbullying etc.

Bots often play a key role in spreading hate speech, as well as fake news, with the purpose of polarizing the public opinion with respect to controversial issues like Brexit or the Catalan referendum. For instance, the authors of a recent study about the 1 Oct 2017 Catalan referendum, showed that in a dataset with 3.6 million tweets, about 23.6% of tweets were produced by bots. The target of these bots were pro independence influencers that were sent negative, emotional and aggressive hateful tweets with hashtags such as #sonunesbesties (i.e. #theyareanimals).

Since 2013 at the PAN Lab at CLEF (https://pan.webis.de/) we have addressed several aspects of author profiling in social media. In 2019 we investigated the feasibility of distinguishing whether the author of a Twitter feed is a bot, while this year we are addressing the problem of profiling those authors that are more likely to spread fake news in Twitter because they did in the past. We aim at identifying possible fake news spreaders as a first step towards preventing fake news from being propagated among online users (fake news aim to polarize the public opinion and may contain hate speech).

In 2021 we specifically aim at addressing the challenging problem of profiling haters in social media in order to monitor abusive language and prevent cases of social exclusion in order to combat, for instance, racism, xenophobia and misoginy. Although we already started addressing the problem of detecting hate speech when targets are immigrants or women at the HatEval shared task in SemEval-2019, and when targets are women also in the Automatic Misogeny Identification tasks at IberEval-2018, Evalita-2018 and Evalita-2020, it was not done from an author profiling perspective. At the end of the keynote I will present some insights in order to stress the importance of monitoring abusive language in social media, for instance, in foreseeing sexual crimes. In fact, previous studies confirmed that a correlation might lay between the yearly per capita rate of rape and the misogynistic language used in Twitter.