SORSE

Towards an NLP Pipeline for Conflict Narrative Detection

Stephen Anning[1], **Dr. George Konstantinidis**[1], **and Dr. Craig Webber**[1]

1 University of Southampton

Software

- SORSE
- Event Website

Category: talks
Published: July 17, 2020

This abstract proposes a talk about PhD research into developing an NLP pipeline for Conflict Narrative Detection. In response to increased incidences of online abuse, a new industry of hate speech detection using NLP has emerged. Accordingly, we tested NLP technologies used by this industry to discover how quantitatively analysing language distorts meaning. We compiled a dataset comprising "Mein Kampf" from Hitler, "War on Terror" texts from George Bush and Osama bin Laden, and in how he advocated for non-violence, speeches from Martin Luther King provide control data. We tested both general-purpose and state-of-the-art sentiment analysis technologies from TextBlob, Google and IBM. Where distinctive results would be expected from a dataset of extremes, our tests show that regardless of technical sophistication, these technologies are unable to distinguish abusive from non-abusive texts. We address this problem with quantitatively analysing language by offering Conflict Narrative Detection as a new approach. Using a series of experiments published on GitHub, participants will learn about developing a sociotechnical pipeline to detect conflict narratives using the spaCy NLP python library. "Conflict narrative" means a narrative produced by an orator who intends to legitimise violence against their outgroup. Accordingly, guiding technical design is the theory of "cultural violence" from Peace Research, which explores processes of violence legitimisation. Detecting a conflict narrative means inferring what cultural violence calls the "Self-Other Gradient". What follows is a hypothesis whereby the steeper the Self-Other Gradient in favour of an orator's ingroup, the more legitimate acts of violence against their outgroup become. To infer this gradient, we move beyond quantitatively analysing language by employing qualitative methods, such as hypernymy. Accordingly, qualitative data produced by the pipeline represent the language patterns used to legitimise violence. Participants will learn how the Self-Other Gradient and these language patterns provide new data for tackling online abuse.