

Proposal: Automatic topic modelling of official communication from Spanish far right deputies during the COVID crisis

Olivier R. Philippe

Introduction

The present proposal focuses on the discussion by the Spanish far right elected representative during the COVID-19 crisis. It explores the topics around which the Twitter discourse from a selected VOX members' accounts revolves around. The topics detection model will be perform using different unsupervised topic modelling methods and manual labelling of the resulting topics distribution for further analysis.

Context

Twitter is often used as an alternative media for health information ??? In case of epidemic and pandemic the uncertainty and fast news can lead to an increase of false information as the verification is not often done by the users and lies and false information spread faster than reliable ones ???.

This phenomenon happened during the H1N1 outbreak in 2009 the Ebola in 2013 and the Zika epidemic in 2015 ???, ???, ???. Naturally the COVID-19 pandemic raises the same issues ???,???

There is a complete descriptive publication on the discussion on Twitter about the COVID-19 and the relation with misinformation ???. However, the COVID-19 is more general than a health issues, considering the novelty, extend and impact of the confinement and health misinformation is only one aspect. Populism and far right in particular are also analysed and their own reaction to the COVID-19 and the spreading of conspiracy theories ???, ??? and the increase of hate messages during this pandemic ???.

In this context of rapid change and fertile context for misinformation, it is important to unveil the ensemble of topics that far right talk about, and not only about the risk posed by health misinformation. This is why topic modelling is an important tool to tackle this issue.

Data and methods

Topic modelling has been used in the context of COVID-19 crisis. Wicke and Bolognesi used the latent Dirichlet allocation LDA method an their existing framework to model discourse on COVID-19 ???, Lwin et al. did a sentiment analysis on the Twitter trends ??? and Haman et al. on the communication by state leaders ???.

A contrario to existing studies using large set of available tweets about the COVID-19 ???, or following keywords of pre-identified issues and hashtags, our data target is a finite set of users, the entire elected deputies elected in 2019 and present on Twitter.

The data collection covers the entire period of 1st January 2020 until the 15th of July 2020, accessed using the REST API from Twitter ???. The method used is a combination of a fine-tuned language model ???, ??? on the dataset and using the created word embeddings with LDA ??? to model the topic distribution. From there, it is possible to compare between different discourses made by elected officials in Spain, and unveil the difference between parties and see if the far right differs from the others parties.

References