Textual analysis of a Twitter corpus during the COVID-19 pandemics

Valerio Astuti, Marta Crispino, Marco Langiulli, Juri Marcucci
Bank of Italy, Directorate General for Economics, Statistics and Research, Italy.

Abstract

Text data gathered from social media are extremely up-to-date and have a great potential value for economic research. At the same time, they pose some challenges, as they require different statistical methods from the ones used for traditional data. The aim of this paper is to give a critical overview of three of the most common techniques used to extract information from text data: topic modelling, word embedding and sentiment analysis. We apply these methodologies to data collected from Twitter during the COVID-19 pandemic to investigate the influence the pandemic had on the Italian Twitter community and to discover the topics most actively discussed on the platform.

Using these techniques of automated textual analysis, we are able to make inferences about the most important subjects covered over time and build real-time daily indicators of the sentiment expressed on this platform.

Keywords: *Text as data, Twitter, Big data, Sentiment, COVID-19, Topic analysis, Word Embedding.*