We Have the Best Words: From the Web-scale Extraction and Attribution of Quotes to Analyzing Negativity in U.S. Political Language

Andreas Spitz

University of Konstanz, 78464 Konstanz, Germany

Abstract

A substantial majority of Americans share the belief that the political discourse in the U.S. has recently become more negative, and more than half of them blame this change on Donald Trump. However, as is often the case in politics, talk is cheap and hard data is difficult to come by. To provide quantitative answers (and distribute blame deservedly) we consider the large-scale extraction and attribution of quotes by politicians for the analysis of political discourse. In the first part of this talk, I introduce Quobert, a transformer-based model that exploits the parallelism in news reporting for the extraction and attribution of quotes from news. Using Quotebank, a comprehensive corpus of 235 million unique quotations that we extracted with Quobert from a decade of news, I then demonstrate how this data can be used to quantify trends in the use of political language. In particular, I will focus on the uptick in negativity in U.S. politicians' language after the end of Obama's tenure, quantify the shifts in language tone, and unravel to whom these shifts could feasibly be attributed.

Short Bio

Andreas Spitz is an assistant professor and head of the Data and Information Mining lab at the University of Konstanz. He holds a PhD in computer science from Heidelberg University and visited the EPFL Data Science lab as a postdoctoral researcher. Andreas' research interests lie at the intersection of information retrieval, natural language processing, computational social science, and complex network analysis. He is particularly interested in graph representations of natural language and how they can be used to efficiently query, visualize, and explore large corpora.