

Assignment 4

Text as Data

2023-10-23

Introduction

In this assignment, you are asked to use topic modelling to investigate manifestos from the manifesto project maintained by [WZB](#). You can either use the UK manifestos we looked at together in class, or collect your own set of manifestos by choosing the country/countries, year/years and party/parties you are interested in. You should produce a report which includes your code, that addresses the following aspects of creating a topic model, making sure to answer the questions below.

This time, you will be assessed not only on whether the code gets the right result, but on how you understand and communicate your understanding of the modelling process and how this can answer your research question. The best research question is one that is interesting *and* answerable, but the most important thing is that the research question is answerable with the methods you choose.

You will also be assessed on the presentation of your results, and on the concision and readability of your code.

1. Data acquisition, description, and preparation

Bring together a dataset from the WZB.

What years, countries and parties are included in the dataset? How many texts do you have for each of these?

Prepare your data for topic modelling by creating a document feature matrix. Describe the choices you make here, and comment on how these might affect your final result.

2. Research question

Describe a research question you want to explore with topic modelling. Comment on how answerable this is with the methods and data at your disposal.

3. Topic model development

Create a topic model using your data. Explain to a non-specialist what the topic model does. Comment on the choices you make here in terms of hyperparameter selection and model choice. How might these affect your results and the ability to answer your research question?

4. Topic model description

Describe the topic model. What topics does it contain? How are these distributed across the data?

5. Answering your research question

Use your topic model to answer your research question by showing plots or statistical results. Discuss the implications of what you find, and any limitations inherent in your approach. Discuss how the work could be improved upon in future research.