Replication of Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data by Pablo Barberá (2015, Political Science)

Isabella Mullen (imullen@ucsd.edu)

2024-10-13

Table of contents

Introduction

This experiment uses machine learning to predict the ideological position of Twitter users based on their social media connections. I chose this paper for the computational reproducibility project because my lab is working on a similar project this year. We are using neural

networks to create a methodology that performs ideal point estimation by combining tabular and non-tabular data sources. Barberá's paper is significant because it was one of the first to use social media data for ideal point estimation and the methods he developed were more precise. I believe that it will be valuable for me to reproduce the findings of this paper to get a better understanding of the model he created and where there is room for improvement.

To replicate this experiment I would need to implement the Bayesian Spatial Following model. I would also need to validate the results using additional data sources. The data used for validation could be the same or updated to include years since publishing to determine if the predictions are still as accurate. This additional validation would ensure that the classifications made by the model were correct. The replication data and code is freely available which addresses the first major challenge. This project will be challenging because I am new to latent variable modeling and it could require significant computational power to execute. Additionally, linking Twitter data to external data sets will require developed computational skills and statistical knowledge. I think this project would be best if worked on with a group.

Here is the link to github repository

Here is the link to original paper

Here is the link to the replication data

Methods

Power Analysis

Original effect size, power analysis for samples to achieve 80%, 90%, 95% power to detect that effect size. Considerations of feasibility for selecting planned sample size.

Planned Sample

Planned sample size and/or termination rule, sampling frame, known demographics if any, preselection rules if any.

Materials

All materials - can quote directly from original article - just put the text in quotations and note that this was followed precisely. Or, quote directly and just point out exceptions to what was described in the original article.

Procedure

Can quote directly from original article - just put the text in quotations and note that this was followed precisely. Or, quote directly and just point out exceptions to what was described in the original article.

Analysis Plan

Can also quote directly, though it is less often spelled out effectively for an analysis strategy section. The key is to report an analysis strategy that is as close to the original - data cleaning rules, data exclusion rules, covariates, etc. - as possible.

Clarify key analysis of interest here You can also pre-specify additional analyses you plan to do.

Differences from Original Study

Explicitly describe known differences in sample, setting, procedure, and analysis plan from original study. The goal, of course, is to minimize those differences, but differences will inevitably occur. Also, note whether such differences are anticipated to make a difference based on claims in the original article or subsequent published research on the conditions for obtaining the effect.

Methods Addendum (Post Data Collection)

You can comment this section out prior to final report with data collection.

Actual Sample

Sample size, demographics, data exclusions based on rules spelled out in analysis plan

Differences from pre-data collection methods plan

Any differences from what was described as the original plan, or "none".

Results

Data preparation

Data preparation following the analysis plan.

Confirmatory analysis

The analyses as specified in the analysis plan.

Side-by-side graph with original graph is ideal here

Exploratory analyses

Any follow-up analyses desired (not required).

Discussion

Summary of Replication Attempt

Open the discussion section with a paragraph summarizing the primary result from the confirmatory analysis and the assessment of whether it replicated, partially replicated, or failed to replicate the original result.

Commentary

Add open-ended commentary (if any) reflecting (a) insights from follow-up exploratory analysis, (b) assessment of the meaning of the replication (or not) - e.g., for a failure to replicate, are the differences between original and present study ones that definitely, plausibly, or are unlikely to have been moderators of the result, and (c) discussion of any objections or challenges raised by the current and original authors about the replication attempt. None of these need to be long.