Codebook

Bottom Up? Top Down? Determinants of Issue-Attention in State Politics

Andreu Casas* Oscar Stuhler† Julia Payson‡
Joshua A. Tucker § Richard Bonneau¶ Jonathan Nagler ‖

*Royal Holloway, University of London, *Department of Politics and International Relations*
†Northwestern University, *Department of Sociology*
‡UCLA, *Department of Political Science*
§New York University: *Department of Politics*
¶Genentech
‖New York University: *Department of Politics*

Contents

Dataset A annotated-tweets-no-text-no-author.csv

Tweets annotated for training the BERT classifiers used in the paper.

Variable	Description
tweet_id	Official ID of the tweet.
dataset	Type of user who created the tweet: state legislator (legislator), legislator follower (leg_follower), or media.
major	Major topic code of the Comparative Agendas Project codebook.
minor	Minor topic code of the Comparative Agendas Project codebook.
tokens_n	Number of words in the tweet.

Dataset B avg-daily-att-per-issue-and-group-<YEAR>.csv

Average daily attention each group included in the analysis devoted to each issue in 2018 OR 2021.

Variable	Description
Issue	Major topic code of the Comparative Agendas Project codebook.
Value	Average daily attention to the issue in 2018/2021 by the Group.
Group	Actor included in the analysis: Members of Congress, National Media, State Legislators, State Media, State Partisans, Random public, or Trump.
overall_mean	Average daily attention across all groups.

Dataset C avg-issue-attention-by-party-grand-means.csv

Average daily Democrats and Republican devoted to each issue in 2018 and 2021.

Variable	Description
Democrat	Average attention to that issue by Democrats.
Republican	AVerage attention to that issue by Republicans.
D_over_R	Democrat / Republican.
R_over_D	Republican / Democrat.
Issue	Major topic code of the Comparative Agendas Project codebook.

Dataset D extended_bills_df_ALL-SIMPLIFIED.csv

Bills introduced in US state legislatures during the period of anlaysis. Source: Legiscan.

Variable	Description
bill_id	Unique identifier given by Legiscan to each bill.
state	Two-letter US state code.
status	Legiscan code indicating the status of the bill.
status_date	Date when latest status update.
title	Official title of the bill.
description	Longer bill description.
first_date	Introduction date of the bill.

Dataset E bills-with-topic-predictions.csv

Topic predictions for the state bills in extended_bills_df_ALL-SIMPLIFIED.csv.

Variable	Description
bill_id	Unique identifier given by Legiscan to each bill.
state	Two-letter US state code.
X0-X21	Probability (0-1) of the bill belonging to each major topic of the Comparative Agendas Project.

Dataset F group-day-issue-level-dataset-01-<YEAR>.csv

Average daily attention devoted to each issue by each group under analysis in 2018 OR 2021. Members of Congress pooled together.

Variable	Description
Date	Day.
national_legislators	Daily attention to the topic specified in IssueState, by members of Congress (logit-transformed proportion).
national_media	Daily attention to to the topic specified in IssueState, by national media (logit-transformed proportion).
state_legislators	Daily attention to the topic specified in IssueState, and by state legislators from the state also specified in IssueState (logit-transformed proportion).
national_media	Daily attention toto the topic specified in IssueState, by national media (logit-transformed proportion).
state_media	Daily attention to the topic specified in IssueState, and by local media from the state also specified in IssueState (logit-transformed proportion).
state_partisans	Daily attention to the topic specified in IssueState, and by followers of state legislators from the state also specified in IssueState (logit-transformed proportion).
state_random_partisans	Daily attention to the topic specified in IssueState, and by random users from the state also specified in IssueState (logit-transformed proportion).
Trump	Daily attention to the topic specified in IssueState, by President Trump (logit-transformed proportion).
IssueState	The issue (major topic of the Comparative Agendas Project) and state for that observation.

Dataset G group-day-issue-level-dataset-02-<YEAR>.csv

Same codebook as in group-day-issue-level-dataset-01-<YEAR>.csv, with the difference that IssueState is split into two variables (Issue and state), that the proportions are not logit transformed, and members of Congress are differentiated by state (instead of being pooled all together).

Dataset H group-day-issue-level-dataset-03-<YEAR>.csv

Same codebook as in group-day-issue-level-dataset-01-<YEAR>.csv, with the difference that members of Congress are differentiated by state (instead of being pooled all together); and members of Congress, state legislators, and state partisans are differentiated by party.

Dataset I state-legislators-mentions-by-partisans-summary.csv

Datatset containing information about how often state partisans mention state legislators in their tweets.

Variable	Description
state	State of the partisan.
party	Party ID of the partisan.
day	The date for the information in that row.
stateleg_mentions	Number of tweets in which at least one state legislator from any state under analysis is mentioned.
stateleg_mentions_dem	Number of tweets in which at least one Democratic state legislator from any state under analysis is mentioned.
stateleg_mentions_rep	Number of tweets in which at least one Republican state legislator from any state under analysis is mentioned.
stateleg_mentions_dem	Number of tweets in which at least one Democratic state legislator from the same state is mentioned.
stateleg_mentions_rep	Number of tweets in which at least one Republican state legislator from the same state is mentioned.
stateleg_mentions_samestate	Number of tweets in which at least one legislator from that state (and from any party) is mentioned
stateleg_mentions_samestate_sameparty	Number of tweets in which at least one legislator from that state and from that party is mentioned

Dataset J state-legislators-network-connections-data.csv

Datatset containing information about what users from other groups under analysis, state legislators follow.

Variable	Description
user_id	User Twitter ID for a state legislator.
user_party	Party ID of the state legislator.
day	The date for the information in that row.
user_state	State of the state legislator.
followed_mcs	List of user IDs of the members of Congress that state legislator follows.
followed_mcs_dem	List of user IDs of the Democratic members of Congress that state legislator follows.
followed_mcs_rep	List of user IDs of the Republican members of Congress that state legislator follows.
followed_mcs_dem_n	Number of Democratic members of Congress that state legislator follows.
followed_mcs_rep	Number of Republican members of Congress that state legislator follows.
followed_mcs_sameparty_n	Number of members of Congress from the same party followed.
followed_mcs_samestate_n	Number of members of Congress from the same state followed.
followed_mcs_samestate_dem_n	Number of Democratic members of Congress from the same state followed.
followed_mcs_samestate_rep_n	Number of Republican members of Congress from the same state followed.
followed_mcs_samestate_sameparty_n	Number of members of Congress from the same state and party followed.
followed_natmedia	List of national media accounts (out of the 4 included in the analysis) the state legislator follows.
followed_natmedia_n	Number of national media accounts (out of the 4 included in the analysis) the state legislator follows.
followed_statemedia	List of state media account (from any state) the state legislator follows.
followed_statemedia_n	Number of state media account (from any state) the state legislator follows.
followed_statemedia_samestate	List of state media account (from the same state) the state legislator follows.
followed_statemedia_samestate_n	Number of state media account (from the same state) the state legislator follows.
followed_partisans_n	Number of state partisans (from any state) followed.
followed_partisans_n	Number of state partisans (from the same state) followed.

Dataset K ml_performance

The naming convention of the files in this directory is the following: <group>-<model-type>-<training-date>-<training-time>.csv, where group indicates whether is the Politician (pol), Media (media) or Partisan (legfol) model; where model-type indicates whether is a bert, roberta, or svm model, and the training date is in YYYY-DD-MM format, and the training time is in HH-MM-SS format. The files contain the following variables:

Variable	Description
epoch	The training epoch.
fold	The training fold.
macro_test	The accuracy based on all test observations.
policy_test	The test accuracy based on policy-relevant tweets.
macro_val	The accuracy based on all validation observations.
policy_val	The validation accuracy based on policy-relevant tweets.
dataset	The data combinationed used for training.
policy_val_f1macro	The validation macro f1score based on policy-relevant tweets.
policy_val_f1weighted	The validation micro f1score based on policy-relevant tweets.