Using Social Media Data to Reveal Patterns of Policy Engagement in State Legislatures

Codebook

1 tweets-only-id-variables.csv

A dataset with the with the collected 2018 tweets for the state legislators under analysis.

Variable Name	Variable Type	Values
user_name	String	Twitter handles of state legislators under analysis
user_id	Numeric	Twitter ID of state legislators under analysis
tweet_id	Numeric	Twitter ID for tweets sent by the state legislators under analysis
date	String	Date in which the tweet was sent (format: YYYY-M-DD)

2 sample-10000-legtweets.csv

A dataset with about 10,000 tweets sent in 2018 by state legislators under analysis, and the topic we predicted the tweets to be about. We use this dataset to explore what are the most distinctive text features for each topic.

Variable Name	Variable Type	Values
tweet_id	Numeric	Twitter ID for tweets in the dataset
user_id	Numeric	Twitter ID of the state legislators who sent the tweet
top_topic	Categorical	The topic into which the tweet has been classified (based on the codebook of the Comparative Agendas Project)
clean_text_cnn	String	Contains the processed text used to generate topic predictions with the CNN model

3 model_data_basic.csv

A dataset with the basic legislator-level covariates included in the regressions reported in the paper.

Variable Name	Variable Type	Values
State	Categorical	The 2-letter abbreviation for the state of the legislator (e.g. AZ)
Chamber	Categorical	The chamber in which the state legislator is serving: either 'House/Assembly' or 'Senate'
District	Categorical	The state district number (or name) the state legislator represents.
Name	String	Full name of the state legislator
Party	Categorical	The party of the state legislator

office.firstElect	String	Date in which the state legislator was first elected for that position
office.lastElect	String	Date in which the state legislator was last elected for that position
committee.count	Numeric	Number of committees in which the state legislator was serving in 2018
mov02	Numeric	Margin of Victory in the last election (the difference points between their voting share and the voting share of the loosing opponent – 100% if the member was uncontested)
ethnicity	Categorical	Ethnicity of the state legislator: 'asian', 'black', 'hispanic', 'white', or 'other_ethn'. These are based on machine learning predictions that used their last name as input (using the 'wru' package in R.)
male	Binary	= 1 if state legislator is male, = 0 if female. This is also based on machine learning predictions using their first name as input (also using the same 'wru' package)
year_first_elected	Categorical	Year in which the representative was first elected to the chamber
seniority	Numeric	Years since first elected
leadership	Numeric	= 1 if state legislator held some leadership position in the chamber in 2018, = 0 otherwise
professional	Categorical	Whether the legislator serves in a legislature considered to be in the group of: 'Least professional', 'Middle-ground', or 'Most professional', based on the [**?**] and data from the *Correlates of State Policy Project*
election18	Binary	Whether there was an mid-term election in 2018 in the legislature in which the state legislator served
legprofscore	Numeric	Legislative professionalization score for the legislature in which the representative served. Source: *Correlates of State Policy Project*
term_limit	Numeric	The number of terms a legislator can serve in that legislature (= 1000 if unlimited terms)
term_length	Numeric	How long a term is in that legislature (in years)
has_term_limits	Binary	Whether the legislature has term limits
last_term	Binary	Whether the legislator is serving the last term
staff	Numeric	Number of staff working for the legislature. Source: https://www.ncsl.org/research/about-state-legislatures/staff-change-chart-1979-1988-1996-2003-2009.aspx
*_sd	Numeric	Standardized version of some of the previous variables, expressed in standard deviation changes.

4 legislators_topic_tweet_count-25june2020.csv

A dataset with information about the number of tweets each legislator sent about each topic in 2018.

Variable Name	Variable Type	Values
`user_id`	Numeric	Twitter ID of the state legislator
`party`	Categorical	The party of the state legislator: 'democrat' or 'republican'
`state`	Categorical	The 2-letter abbreviation for the state of the legislator (e.g. AZ)
`Chamber`	Categorical	The chamber in which the state legislator is serving: either 'House/Assembly' or 'Senate'
`X*`	Numeric	Several variables indicating how many tweets on those topics the legislator sent in 2018. Here the topic each code represents: X0 (No policy issue), X1 (Economy), X2 (Law and Crime), X3 (Defense), X4 (Technology), X5 (Foreign Trade), X6 (Intl. Affairs), X7 (Govt. Operations), X8 (Public Lands), X9 (Gun Control), X10 (Civil Rights), X12 (Healthcare), X13 (Agriculture), X14 (Labor), X15 (Education), X16 (Environment), X17 (Energy), X18 (Immigration), X19 (Transportation), X20 (Social Welfare), X21 (Housing), X25 (Domestic Commerce).

5 final-LEG-model-acc-bytopic-VALSET.csv

A dataset used to assess the accuracy of the model by topic.

Variable Name	Variable Type	Values
`X_val`	String	The processed text used to generate topic predictions with the CNN model
`tokens`	Numeric	The number of tokens in the processed text
`Y_val`	Categorical	The manually annotated topic code for this tweet (gold standard)
`ccn_prediction`	Categorical	The topic code predicted by the CNN model

6 memberscongress_topic_tweet_count-25june2020.csv

A dataset with information about the number of tweets members of Congress sent about each topic in 2018.

Variable Name	Variable Type	Values
`user_id`	Numeric	Twitter ID of the members of Congress
`party`	Categorical	The party of the member of Congress: 'democrat' or 'republican'
`state`	Categorical	The 2-letter abbreviation for the state the members of Congress represents (e.g. AZ)
`chamber`	Categorical	The chamber in which the members of Congress is serving: either 'house' or 'senate'
`X*`	Numeric	Several variables indicating how many tweets on those topics the members of Congress sent in 2018. Here the topic each code represents: X0 (No policy issue), X1 (Economy), X2 (Law and Crime), X3 (Defense), X4 (Technology), X5 (Foreign Trade), X6 (Intl. Affairs), X7 (Govt. Operations), X8 (Public Lands), X9 (Gun Control), X10 (Civil Rights), X12 (Healthcare), X13 (Agriculture), X14 (Labor), X15 (Education), X16 (Environment), X17 (Energy), X18 (Immigration), X19 (Transportation), X20 (Social Welfare), X21 (Housing), X25 (Domestic Commerce).

7 committee_topic.csv

A dataset with information about whether a state legislator served in a committee about each of the topics under study.

Variable Name	Variable Type	Values
`user_id`	Numeric	Twitter ID of the state legislator
<`remaining`>	Binary	Whether the state legislator served on a committee on the topic indicated in the name of the column

8 state-legislatures-when-in-session-2018.csv

A dataset with information about when the state legislatures of the representatives under analysis convene and adjourn.

Variable Name	Variable Type	Values
`state`	Categorical	The 2-letter abbreviation for the state (e.g. AZ)
`convenes`	String	Date in which the state legislatures starts the session (format: M-DD-YYYY)
`adjourns`	String	Date in which the state legislatures ends the session (format: M-DD-YYYY)

9 state-leg-indiv-level-meta-twitter-basics-2july2020.csv

A dataset with further information about each individual state legislator under analysis.

Variable Name	Variable Type	Values
`State`	Categorical	The 2-letter abbreviation for the state (e.g. AZ)
`Chamber`	Categorical	The chamber in which the state legislator is serving: either 'House/Assembly' or 'Senate'
`District`	Categorical	The state district number (or name) the state legislator represents.
`Name`	String	Full name of the state legislator
`Twitter`	String	Twitter screen name of the state legislator
`Party`	Categorical	The party of the state legislator
`updated2019`	Binary	If the legislator was elected in the midterm 2018 election (= 1, 0 otherwise)
`user_id`	Numeric	Twitter ID of state legislators under analysis
`tweets_n`	Numeric	Number of tweets sent in 2018