Political Mail Formatting and Voter Turnout

Marcy Shieh - Blake Reynolds

University of Wisconsin - Madison

Campaign Mali Format

- **Question** Does the presentation of campaign mail influence voter turnout?
- **Introduction** This project seeks to investigate whether attributes of a campaign mailer, such as the size of logos, the style of texts, the presence of blockquotes and endorsements, and information about voting locations and absentee voting can help predict whether a targeted voter or members of the targeted voters' household turn out to vote. Furthermore, we compare the performance of a model with only image-level attributes of campaign mail to models that include voter characteristics.

Voter Data

- Our data contains validated voter turnout data from the 2018 Republican Primary and General Elections in Texas.
- We look at both targeted voters and targeted households. Targeted voters are the individuals that the campaign wants to motivate to turn out to vote for the candidate, and targeted households include members who live in the same household as the targeted voters.
- Voter characteristics in the dataset include gender, age, ethnicity, marital status, and particular interests among many others. Each of these characteristics are individual voter-level demographic variables and come from a mix of commercial voter file products.

Target Primary Election Voters	474,726
Target General Election Voters	96,713
Household Primary Election Voters	972,743
Household General Election Voters	173,545

Fig. 1: Total Universe size of voters who received mail

Methods

- We randomly split the turnout datasets for targeted voters and targeted households for primary and general elections into separate training and testing datasets, placing 80% of the data into the training dataset and the remaining 20% of the data into the testing dataset.
- Models capture targeted voters in primary elections, targeted voters in general elections, targeted households in primary elections, and targeted households in general elections.
- For each type of model formula, we set the train control values for all the models within the type. To do this, we use the cross-validation method with 10 folds and we look at sensitivity, specificity, and the area under the ROC curve as well as compute class probabilities for the model and save all hold-out predictions for each resample.
- We then fit our predictive models using the model formulas. We use training data from the four types of datasets. We specify the parameters, with ROC as the summary metric that we want to use to obtain the optimal model and GLM as our classification method. We then extract the fitted values from the optimal model. From there, we standardize the input data evaluate the performance of the predictive models by looking at the false positive rates and true positive rates.
- Using a ROC curve, we can examine two different types of targets (voter and household) and two different types of elections (primary and general).

Mailpiece Data

- We use a total of 309 mailpiece in our dataset. Each mailpiece is a side–front or back– of a print advertisement that the targeted voters and targeted households received through postal mail.
- We hand-code attributes, use image processing tools to identify additional attributes, and produce principal components (PC) for each mailpiece.
- Each PCA is generated on a single unpixelated image and limited to 25 components. This is enough to explain greater than 85% of the cumulative variance in each image. The x-value, rotation, standard deviation, and center of each PC were saved and collapsed to their mean for each image file.
- We extract hue, saturation, and light values on each image at a 50x50 pixelated scale.

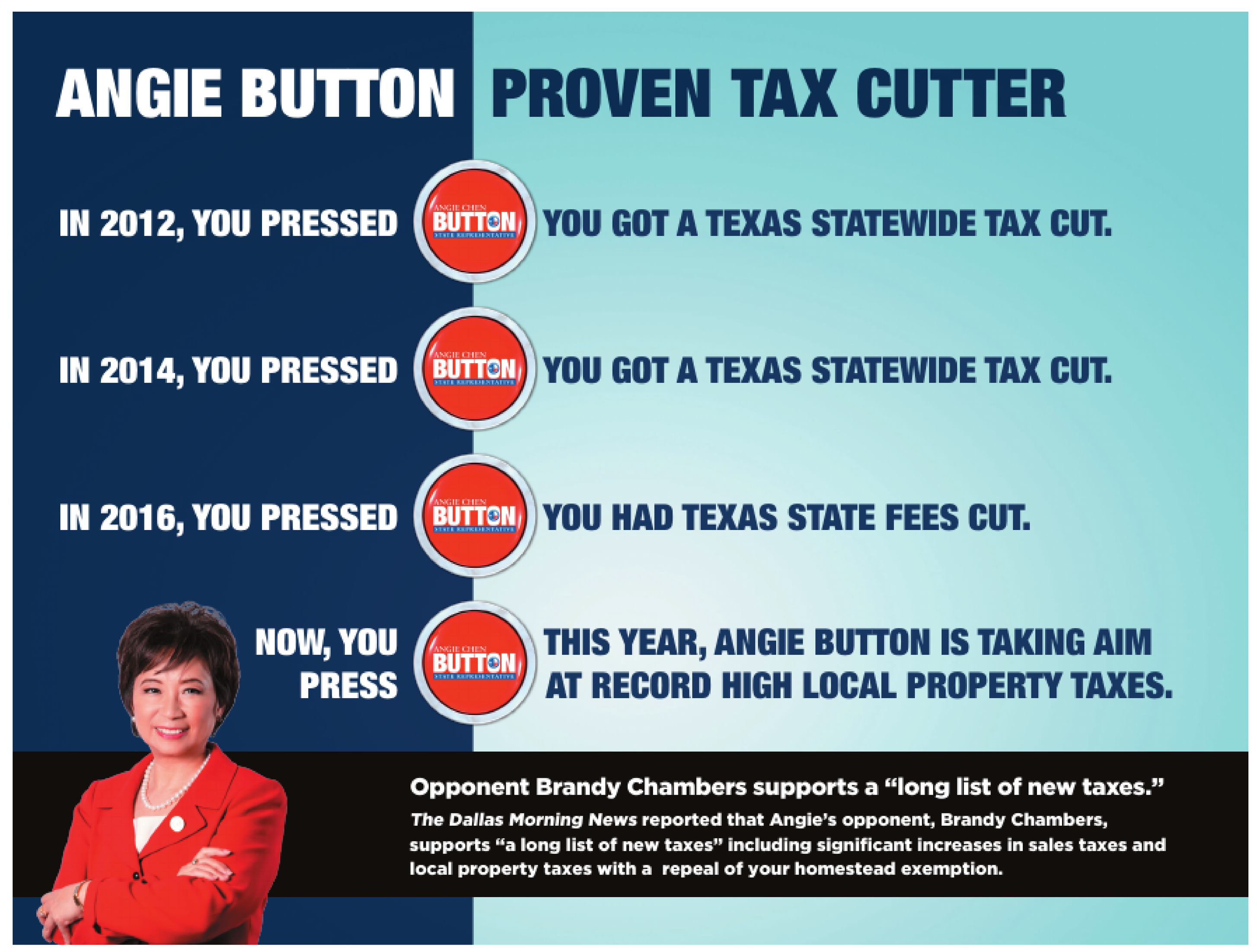

Fig. 2: Example of mailpiece within the dataset. This image represents one mailpiece (i.e., a side of a print ad).

Test Model

Hand-coded Variables Characteristics The first type of model formula predicts turnout of targets using hand-coded variables (δ) and the characteristics (τ) of the targets.

$$Y_i = \beta_0 + \beta_\delta X_{i_t} + \beta_\tau X_{i_t} + \epsilon_{i_t}$$

PC of Mailpieces Characteristics The second type of model formula predicts turnout of targets using PC of mailpieces (ρ) and the characteristics (τ) of the targets.

$$Y_i = \beta_0 + \beta_\rho X_{i_t} + \beta_\tau X_{i_t} + \epsilon_{i_t}$$

Hand-coded Variables and PC of Mailpieces The third type of model formula predicts turnout of targets using both hand-coded variables (δ) and PC of mailpieces (ρ) without the characteristics of the targets.

$$Y_i = \beta_0 + \beta_\delta X_{i_t} + \beta_\rho X_{i_t} + \epsilon_{i_t}$$

Characteristics The fourth type of model formula is the nominal turnout model of targets using the characteristics of the targets.

Results

Model	Election	Variables	Accuracy	NIR	95% CI
1	Primary	Char. + Image	0.86	0.51	(0.857, 0.862)
9	Primary	Image Only	0.60	0.51	(0.594, 0.602)
A	Primary	Nominal	0.85	0.51	(0.853, 0.857)
2	General	Char. + Image	0.84	0.74	(0.835, 0.845)
10	General	Image Only	0.78	0.74	(0.778, 0.789)
B	General	Nominal	0.84	0.74	(0.831, 0.842) height

Fig. 3: Table of Selected Results (Targets Only)

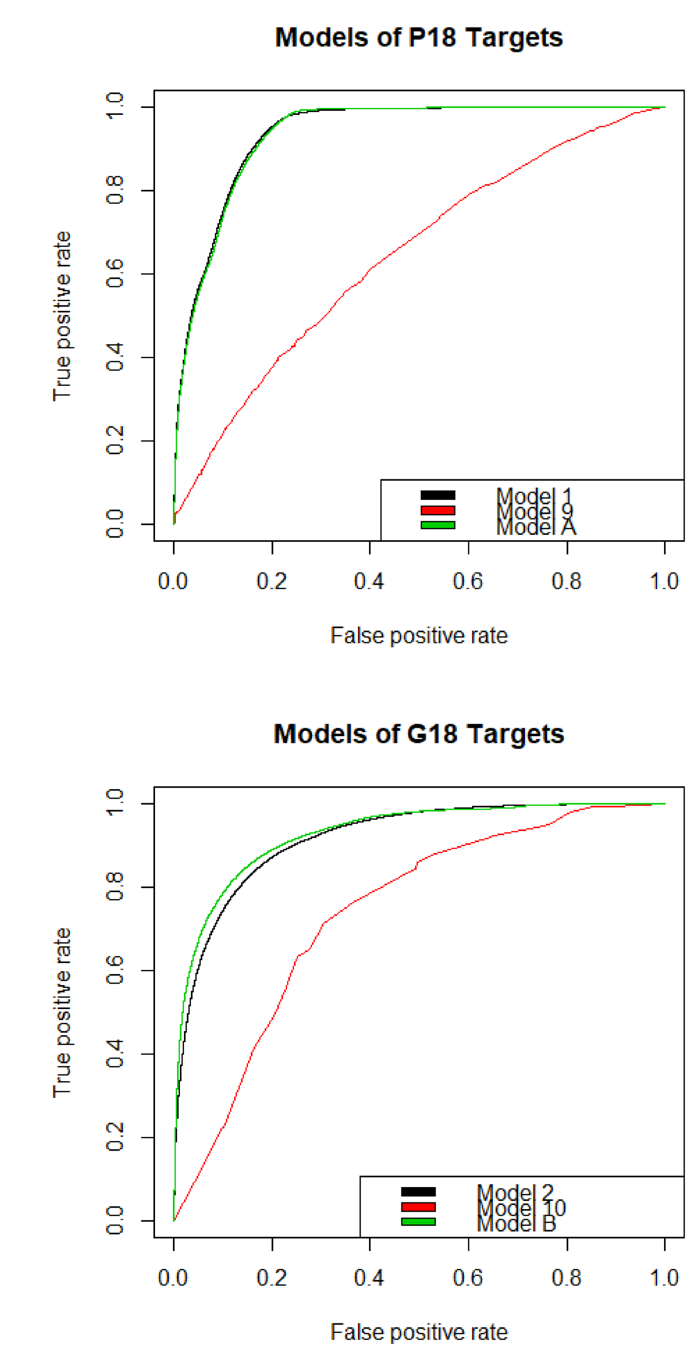

Fig. 4: ROC Plots of Selected Models from Fig. 3.

Future Prospects

- Consider using RMSE instead of sensitivity and specificity statistics.
- Identify attributes in mailers that have the most predictive information.
- Frame the results that would be useful for a public policy audience.
- Address the substantive problem of multiple mailers.
- Work with a campaign consulting firm to design a field experiment with custom ads that include/exclude certain attributes targeted at voters.

Acknowledgements

We thank Adeline Lo and our colleagues in our machine learning seminar for their helpful guidance and feedback.