Machine learning methods for political campaigns

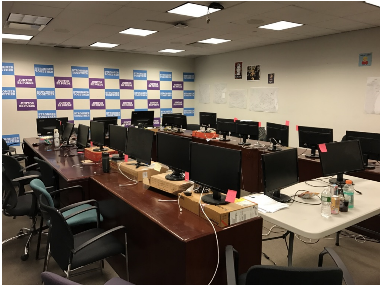
Valerie Bradley

University of Oxford, Department of Statistics

15 November 2019

Background

My background



Political campaign analytics overview

How are analytics used in campaigns?

Goal: win more votes / seats / power than the other candidate(s)

How are analytics used in campaigns?

Goal: win more votes / seats / power than the other candidate(s)
Key questions for a campaign:

How are analytics used in campaigns?

Goal: win more votes / seats / power than the other candidate(s)
Key questions for a campaign:

1. **Current state of the race**. If we did nothing, who would win?
2. **Develop "path to victory"**. What do we have to do to move from current state of the race to winning state? Who do we have to persuade v. who do we have to make sure turns out to vote?

How are analytics used in campaigns?

Goal: win more votes / seats / power than the other candidate(s)
Key questions for a campaign:

1. **Current state of the race**. If we did nothing, who would win?
2. **Develop "path to victory"**. What do we have to do to move from current state of the race to winning state? Who do we have to persuade v. who do we have to make sure turns out to vote?
3. **Optimization**. How do we spend our budget (eg. across modes of outreach like TV, digital, mail, etc., or across targets within each mode) to maximize our impact on key demograpic groups / regions?

How are analytics used in campaigns?

Goal: win more votes / seats / power than the other candidate(s)
Key questions for a campaign:

1. **Current state of the race**. If we did nothing, who would win?
2. **Develop "path to victory"**. What do we have to do to move from current state of the race to winning state? Who do we have to persuade v. who do we have to make sure turns out to vote?
3. **Optimization**. How do we spend our budget (eg. across modes of outreach like TV, digital, mail, etc., or across targets within each mode) to maximize our impact on key demograpic groups / regions?
4. **Tracking**. Are we moving people? Are we losing ground? Are we hitting goals?

The basic analytics products

The basic analytics products

1. State-of-the-race models:
   1. **Turnout score**, $\hat{p}(T = 1|x_i)$
   2. **Support score**, $\hat{p}(Y = 1|x_i)$

The basic analytics products

1. State-of-the-race models:
    1. **Turnout score**, $\hat{p}(T = 1|x_i)$
    2. **Support score**, $\hat{p}(Y = 1|x_i)$
2. Behavior change models:
    1. **Persuasion score**,
       $\hat{p}(P = 1|x_i, Z = 1) = \hat{p}(Y = 1|x_i, Z = 1) - \hat{p}(Y = 1|x_i, Z = 0)$,
       where Z is a treatment indicator
    2. **Get Out The Vote (GOTV) score** , $\hat{p}(T = 1|x_i, C = 1)$
       where $C = 1$ if contacted by the campaign

The basic analytics products

1. State-of-the-race models:
    1. **Turnout score**, $\hat{p}(T = 1|x_i)$
    2. **Support score**, $\hat{p}(Y = 1|x_i)$
2. Behavior change models:
    1. **Persuasion score**,
    $\hat{p}(P = 1|x_i, Z = 1) = \hat{p}(Y = 1|x_i, Z = 1) - \hat{p}(Y = 1|x_i, Z = 0)$,
    where Z is a treatment indicator
    2. **Get Out The Vote (GOTV) score** , $\hat{p}(T = 1|x_i, C = 1)$
    where $C = 1$ if contacted by the campaign
3. Mode contactability / impact models
    1. **Poll response rate models**: $\hat{p}(R = 1|x_i, [\text{attempted}])$
    2. **Impact models**: $\hat{p}(I = 1|x_i, [\text{recieve an ad}])$

The data

Main sources of data:

The data

Main sources of data:

- **Voterfile**: List of all registered voters in the country (historical turnout data, registration address and phone number, age, gender, etc.)

The data

Main sources of data:

- **Voterfile**: List of all registered voters in the country (historical turnout data, registration address and phone number, age, gender, etc.)
- **Consumer and census data**: appended to the (publicly available) voterfile

The data

Main sources of data:

- **Voterfile**: List of all registered voters in the country (historical turnout data, registration address and phone number, age, gender, etc.)
- **Consumer and census data**: appended to the (publicly available) voterfile
- **Contact history data**: voter interaction data collected by campaigns over many election cycles

The data

Main sources of data:

- **Voterfile**: List of all registered voters in the country (historical turnout data, registration address and phone number, age, gender, etc.)
- **Consumer and census data**: appended to the (publicly available) voterfile
- **Contact history data**: voter interaction data collected by campaigns over many election cycles
- **Historical results and scores**: most granular unit of results is precinct-level (approx. 1000-3000 registered voters), adjust previous scores to actual election results

The data

Main sources of data:

- **Voterfile**: List of all registered voters in the country (historical turnout data, registration address and phone number, age, gender, etc.)
- **Consumer and census data**: appended to the (publicly available) voterfile
- **Contact history data**: voter interaction data collected by campaigns over many election cycles
- **Historical results and scores**: most granular unit of results is precinct-level (approx. 1000-3000 registered voters), adjust previous scores to actual election results
- **Survey data**: data collected in-cycle on the current state of the race, either over the phone or online; different types of polls (modeling, tracking, message testing)

Common problems

Some common problems in political campaign analytics:

Common problems

Some common problems in political campaign analytics:

- We never observe the outcome we try to predict (vote choice at the individual-level)

Common problems

Some common problems in political campaign analytics:

- We never observe the outcome we try to predict (vote choice at the individual-level)
- Quality of polls and models judged by accuracy of e-day prediction, but the outcome used in training is observed at a different moment in time (and opinions change!)

Common problems

Some common problems in political campaign analytics:

- We never observe the outcome we try to predict (vote choice at the individual-level)
- Quality of polls and models judged by accuracy of e-day prediction, but the outcome used in training is observed at a different moment in time (and opinions change!)
- Contact history data is biased (collection is targeted using modeled scores)

Common problems

Some common problems in political campaign analytics:

- We never observe the outcome we try to predict (vote choice at the individual-level)
- Quality of polls and models judged by accuracy of e-day prediction, but the outcome used in training is observed at a different moment in time (and opinions change!)
- Contact history data is biased (collection is targeted using modeled scores)
- Survey data is really expensive to collect (a single response costs 10-20 USD); people don't respond

Common problems

Some common problems in political campaign analytics:

- We never observe the outcome we try to predict (vote choice at the individual-level)
- Quality of polls and models judged by accuracy of e-day prediction, but the outcome used in training is observed at a different moment in time (and opinions change!)
- Contact history data is biased (collection is targeted using modeled scores)
- Survey data is really expensive to collect (a single response costs 10-20 USD); people don't respond
- Standard supervised learning techniques require an outcome (eg. survey response) matched 1:1 to covariate data (eg. voterfile data), so the survey modes we can use are limited; harder with new data privacy restrictions

Common problems

Some common problems in political campaign analytics:

- We never observe the outcome we try to predict (vote choice at the individual-level)
- Quality of polls and models judged by accuracy of e-day prediction, but the outcome used in training is observed at a different moment in time (and opinions change!)
- Contact history data is biased (collection is targeted using modeled scores)
- **Survey data is really expensive to collect (a single response costs 10-20 USD); people don't respond**
- **Standard supervised learning techniques require an outcome (eg. survey response) matched 1:1 to covariate data (eg. voterfile data), so the survey modes we can use are limited; harder with new data privacy restrictions**

Project 1: Methods for selection bias

Methods for selection bias

Selection bias: "preferential exclusion of units from samples" with unknown probability (Bareinboim, 2014). Interested in $\mathrm{P}(y)$, but only observe $\mathrm{P}(y|S=1)$ and $Y \not\perp\!\!\!\perp S$

Methods for selection bias

Selection bias: "preferential exclusion of units from samples" with unknown probability (Bareinboim, 2014). Interested in $\mathrm{P}(y)$, but only observe $\mathrm{P}(y|S=1)$ and $Y \not\perp\!\!\!\perp S$

What can we do? Try to find a set of auxilary variables $\mathbf{Z}$ s.t. $Y \perp\!\!\!\perp S|\mathbf{Z}$

Methods for selection bias

Selection bias: "preferential exclusion of units from samples" with unknown probability (Bareinboim, 2014). Interested in $\mathrm{P}(y)$, but only observe $\mathrm{P}(y|S=1)$ and $Y \not\perp\!\!\!\perp S$

What can we do? Try to find a set of auxilary variables $\mathbf{Z}$ s.t. $Y \perp\!\!\!\perp S|\mathbf{Z}$
Then, we can recover $\mathrm{P}(y)$ with:

$$\mathrm{P}(y) = \sum_{\mathbf{z}} \mathrm{P}(y|\mathbf{z}, S=1)\mathrm{P}(\mathbf{z}) \tag{1}$$

Methods for selection bias

Two main families of methods:

Methods for selection bias

Two main families of methods:

- **Regression methods**: estimate $\mathrm{P}(y|\mathbf{z}, S = 1)$ then project regression model on population distribution $P(\mathbf{z})$
  - MRP (Mister P) is a **really** popular method in political science for this

Methods for selection bias

Two main families of methods:

- **Regression methods**: estimate $\mathrm{P}(y|\mathbf{z}, S=1)$ then project regression model on population distribution $P(\mathbf{z})$
  - MRP (Mister P) is a **really** popular method in political science for this
- **Inverse probability weighting (IPW)**: Estimate the probability of selection $\mathrm{P}(S=1|\mathbf{z})$, then assign observations weights of $1/\hat{\mathrm{P}}(S=1|\mathbf{z})$ and use to calculate weighted estimate of $\mathrm{P}(y)$.

Methods for selection bias

Two main families of methods:

- **Regression methods**: estimate $\mathrm{P}(y|\mathbf{z}, S=1)$ then project regression model on population distribution $P(\mathbf{z})$
  - MRP (Mister P) is a **really** popular method in political science for this
- **Inverse probability weighting (IPW)**: Estimate the probability of selection $\mathrm{P}(S=1|\mathbf{z})$, then assign observations weights of $1/\hat{\mathrm{P}}(S=1|\mathbf{z})$ and use to calculate weighted estimate of $\mathrm{P}(y)$.
  This works because we can re-express Eq. 1:

$$\mathrm{P}(\mathbf{y}) = \sum_{\mathbf{z}} \mathrm{P}(\mathbf{y}|\mathbf{z}, S=1)\mathrm{P}(\mathbf{z}) = \sum_{\mathbf{z}} \mathrm{P}(\mathbf{y}, \mathbf{z}|S=1)\frac{\mathrm{P}(S=1)}{\mathrm{P}(S=1|\mathbf{z})} \quad (2)$$

Methods for selection bias

Two main families of methods:

- **Regression methods**: estimate $\mathrm{P}(y|\mathbf{z}, S = 1)$ then project regression model on population distribution $P(\mathbf{z})$
  - MRP (Mister P) is a **really** popular method in political science for this
- **Inverse probability weighting (IPW)**: Estimate the probability of selection $\mathrm{P}(S = 1|\mathbf{z})$, then assign observations weights of $1/\hat{\mathrm{P}}(S = 1|\mathbf{z})$ and use to calculate weighted estimate of $\mathrm{P}(y)$.
  This works because we can re-express Eq. 1:

$$\mathrm{P}(\mathbf{y}) = \sum_{\mathbf{Z}} \mathrm{P}(\mathbf{y}|\mathbf{z}, S = 1)\mathrm{P}(\mathbf{z}) = \sum_{\mathbf{Z}} \mathrm{P}(\mathbf{y}, \mathbf{z}|S = 1)\frac{\mathrm{P}(S = 1)}{\mathrm{P}(S = 1|\mathbf{z})} \tag{2}$$

Problem: What if $\mathbf{Z}$ is large? Or contains continuous variables?

Methods for selection bias

Two main families of methods:

- **Regression methods**: estimate $\mathrm{P}(y|\mathbf{z}, S=1)$ then project regression model on population distribution $P(\mathbf{z})$
  - MRP (Mister P) is a **really** popular method in political science for this
- **Inverse probability weighting (IPW)**: Estimate the probability of selection $\mathrm{P}(S=1|\mathbf{z})$, then assign observations weights of $1/\hat{\mathrm{P}}(S=1|\mathbf{z})$ and use to calculate weighted estimate of $\mathrm{P}(y)$.
  This works because we can re-express Eq. 1:

$$\mathrm{P}(\mathbf{y}) = \sum_{\mathbf{Z}} \mathrm{P}(\mathbf{y}|\mathbf{z}, S=1)\mathrm{P}(\mathbf{z}) = \sum_{\mathbf{Z}} \mathrm{P}(\mathbf{y}, \mathbf{z}|S=1)\frac{\mathrm{P}(S=1)}{\mathrm{P}(S=1|\mathbf{z})} \quad (2)$$

Problem: What if $\mathbf{Z}$ is large? Or contains continuous variables?

Which to use? Regression is flexible, but outcome dependent

Estimating $\text{P}(S = 1|\mathbf{Z})$

Things to think about when comparing methods for estimating $\text{P}(S = 1|\mathbf{Z})$:

1. Weighted **marginal distribution** of $\mathbf{Z}$ should match that of the population
2. Avoid extreme weights that inflate the **variance** of estimators (have to account for the additional uncertainty of weights)
3. **Computational complexity** of the method (weighting should be FAST, otherwise you should just build a model for Y)
4. How well does the method account for **interactions** between elements of $\mathbf{Z}$

MANY IPW methods (I)

Classic weighting methods:

1. *Post-stratification*: adjust to the joint distribution of $\mathbf{Z}$ (exactly Eq 2)
2. *Raking*: iteratively adjust the marginal distributions of elements in $\mathbf{Z}$
3. *Calibration*: raking, but can handle continuous variables (use population totals of continuous variables as constraints)

MANY IPW methods (I)

Classic weighting methods:

1. *Post-stratification*: adjust to the joint distribution of $\mathbf{Z}$ (exactly Eq 2)
2. *Raking*: iteratively adjust the marginal distributions of elements in $\mathbf{Z}$
3. *Calibration*: raking, but can handle continuous variables (use population totals of continuous variables as constraints)

Less-common methods:

4. *Logit*: estimate the $\mathrm{P}(S = 1|\mathbf{z})$ directly with regularized regression
5. *LASSO*: use a LASSO to select important subset of $\mathbf{Z}$, then rake to subset

MANY IPW methods (II)

New methods:

6. *BART + raking*: use a Bayesian Additive Regression Tree (BART) to estimate the probability of selection, then rake such that key marginal distributions match those of the population

MANY IPW methods (II)

New methods:

6. *BART + raking*: use a Bayesian Additive Regression Tree (BART) to estimate the probability of selection, then rake such that key marginal distributions match those of the population
7. *Kernel Mean Matching (KMM)*: Estimate weights $\mathbf{w}$ that minimize distance in RKHS between sample (s) and population (p) covariate distributions

$$\underset{\mathbf{w}}{\operatorname{argmin}} \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} w_i \phi(z_i^s) - \frac{1}{n_p} \sum_{i=1}^{n_p} \phi(z_i^p) \right\|$$

$$\text{subject to } \mathbf{w} \in [0, B] \text{ and } \left| \sum_{i=1}^{n_s} w_i - n_s \right| \leq n_s \epsilon$$

MANY IPW methods (II)

New methods:

6. *BART + raking*: use a Bayesian Additive Regression Tree (BART) to estimate the probability of selection, then rake such that key marginal distributions match those of the population

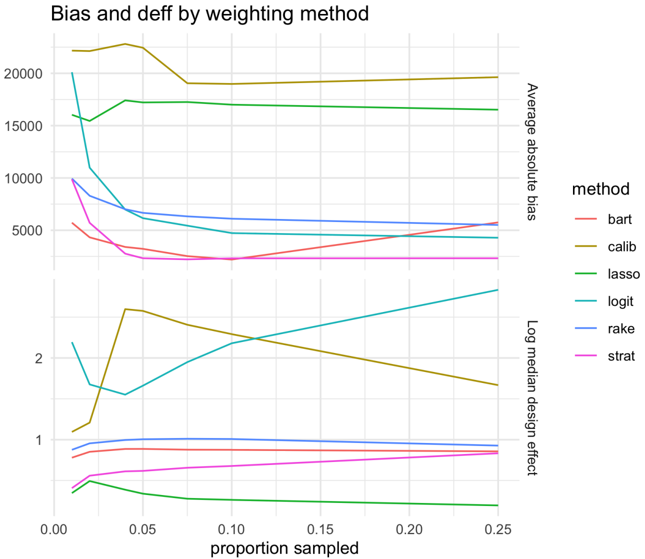
7. *Kernel Mean Matching (KMM)*: Estimate weights **w** that minimize distance in RKHS between sample (s) and population (p) covariate distributions

$$\underset{\mathbf{w}}{\operatorname{argmin}} \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} w_i \phi(z_i^s) - \frac{1}{n_p} \sum_{i=1}^{n_p} \phi(z_i^p) \right\|$$

$$\text{subject to } \mathbf{w} \in [0, B] \text{ and } \left| \sum_{i=1}^{n_s} w_i - n_s \right| \leq n_s \epsilon$$

8. *Post-stratification with* `k-means++`: Use `k-means++` to define strata for post-stratification

Some results



Project 2: Distribution regression for ecological inference

Some familiar faces

Understanding the 2016 US Presidential Election using ecological inference and distribution regression with census microdata

Seth Flaxman (Department of Statistics, Oxford),*
Dougal J. Sutherland (Gatsby Computational Neuroscience Unit, University College London),
Yu-Xiang Wang (Machine Learning Department, Carnegie Mellon University),
and Yee Whye Teh (Department of Statistics, Oxford)

November 14, 2016

Variational Learning on Aggregate Outputs with Gaussian Processes

Ho Chung Leon Law* University of Oxford
Dino Sejdinovic* University of Oxford
Ewan Cameron† University of Oxford
Tim CD Lucas† University of Oxford
Seth Flaxman‡ Imperial College London
Katherine Battle† University Of Oxford
Kenji Fukumizu§ Institute of Statistical Mathematics

Bayesian Approaches to Distribution Regression

Ho Chung Leon Law* University of Oxford ho.law@spc.ox.ac.uk
Dougal J. Sutherland* University College London dougal@gmail.com
Dino Sejdinovic University of Oxford dino.sejdinovic@stats.ox.ac.uk
Seth Flaxman Imperial College London s.flaxman@imperial.ac.uk

Some familiar faces

Understanding the 2016 US Presidential Election using ecological inference and distribution regression with census microdata

Seth Flaxman (Department of Statistics, Oxford)*,
Dougal J. Sutherland (Gatsby Computational Neuroscience Unit, University College London),
Yu-Xiang Wang (Machine Learning Department, Carnegie Mellon University),
and Yee Whye Teh (Department of Statistics, Oxford)

November 14, 2016

Variational Learning on Aggregate Outputs with Gaussian Processes

Ho Chung Leon Law*, University of Oxford
Dino Sejdinovic*, University of Oxford
Ewan Cameron†, University of Oxford
Tim CD Lucas†, University of Oxford
Seth Flaxman‡, Imperial College London
Katherine Battle†, University Of Oxford
Kenji Fukumizu§, Institute of Statistical Mathematics

Bayesian Approaches to Distribution Regression

Ho Chung Leon Law*, University of Oxford, ho.law@spc.ox.ac.uk
Dougal J. Sutherland*, University College London, dougal@gmail.com
Dino Sejdinovic, University of Oxford, dino.sejdinovic@stats.ox.ac.uk
Seth Flaxman, Imperial College London, s.flaxman@imperial.ac.uk

- **Typical supervised learning**: observe $\{\mathbf{x}_i, y_i\}_{i=1}^{n}$, and learn $y_i = f(\mathbf{x}_i)$

Some familiar faces

Understanding the 2016 US Presidential Election using ecological inference and distribution regression with census microdata

Seth Flaxman (Department of Statistics, Oxford)*,
Dougal J. Sutherland (Gatsby Computational Neuroscience Unit, University College London),
Yu-Xiang Wang (Machine Learning Department, Carnegie Mellon University),
and Yee Whye Teh (Department of Statistics, Oxford)

November 14, 2016

Variational Learning on Aggregate Outputs with Gaussian Processes

Ho Chung Leon Law* University of Oxford
Dino Sejdinovic* University of Oxford
Ewan Cameron† University of Oxford
Tim CD Lucas† University of Oxford
Seth Flaxman‡ Imperial College London
Katherine Battle† University Of Oxford
Kenji Fukumizu§ Institute of Statistical Mathematics

Bayesian Approaches to Distribution Regression

Ho Chung Leon Law* University of Oxford ho.law@spc.ox.ac.uk
Dougal J. Sutherland* University College London dougal@gmail.com
Dino Sejdinovic University of Oxford dino.sejdinovic@stats.ox.ac.uk
Seth Flaxman Imperial College London s.flaxman@imperial.ac.uk

- **Typical supervised learning**: observe $\{\mathbf{x}_i, y_i\}_{i=1}^n$, and learn $y_i = f(\mathbf{x}_i)$
- **Distribution regression**:
  - observe $\left(\{\mathbf{x}_i^1\}_{i=1}^{N_1}, y_1\right), \ldots, \left(\{\mathbf{x}_i^B\}_{i=1}^{N_B}, y_B\right)$ where $\mathbf{x}_i^j \sim \mathrm{P}_j$
  - learn $y_j = f(\mu_{\mathrm{P}_j})$, where μ_{P_j} is the kernel mean embedding of P_j
  - estimate μ_{P} with the empirical kernel mean embedding $\hat{\mu}_{\mathrm{P}_j} = \frac{1}{N_j}\sum_{k=1}^{N_j} k(\cdot, x_i^j)$

A slightly different setting

Recall: common problem in campaign analytics that standard techniques require survey data matched to the voterfile, which is expensive and hard to do with privacy restrictions. Also discarding unmatched data introduces (more) bias.

	Y	X_{svy}	X_{both}	X_{vf}
Unmatched (U)	Y^U	X_{svy}^U	X_{both}^U	X_{vf}^U
Matched (M)	Y^M	X_{svy}^M	X_{both}^M	X_{vf}^M
Voterfile (V)	Y^V	X_{svy}^V	X_{both}^V	X_{vf}^V

Figure 1: Diagram of observed and unobserved data. Blue shading represents data that is observed, yellow represents the target for prediction and white is unobserved. Area represents approximate relative size of subsets of data; rows are observations and columns are types of covariates.

Approach 1

Key difference between this setting and standard distribution regression setting, we observe individual y_i, we just can't match them to covariates.

Idea: Aggregate y_i to move to the traditional distribution regression setting

Approach 1

Key difference between this setting and standard distribution regression setting, we observe individual y_i, we just can't match them to covariates.

Idea: Aggregate y_i to move to the traditional distribution regression setting

1. Define B bags with X_{both} using `k-means++`
2. Aggregate y_i within those bags (mean or frequency counts)
3. Embed bagged (X_{both}, X_{vf}) in RKHS and calculate mean embedding
4. Perform reglarized regression of aggregated $\bar{y}_i$ on mean embeddings

Approach 1

Key difference between this setting and standard distribution regression setting, we observe individual y_i, we just can't match them to covariates.

Idea: Aggregate y_i to move to the traditional distribution regression setting

1. Define B bags with X_{both} using `k-means++`
2. Aggregate y_i within those bags (mean or frequency counts)
3. Embed bagged (X_{both}, X_{vf}) in RKHS and calculate mean embedding
4. Perform reglarized regression of aggregated $\bar{y}_i$ on mean embeddings

Variations:

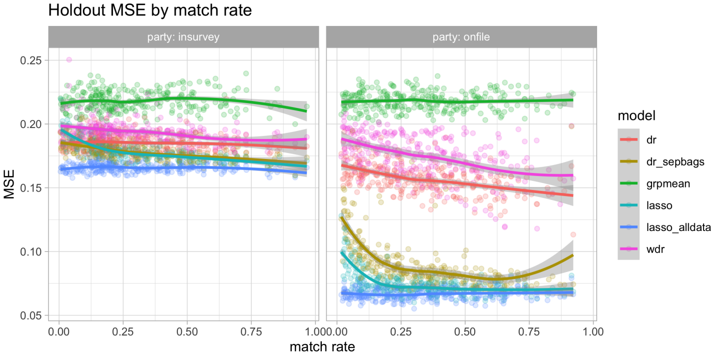
1. Weighted aggregation of y_i using kernel mean matching (KMM)
2. Only aggregate unmatched y_i (avoid information loss from aggregation)

Approach 1 - Simulation results

Data: 5,259 responses from 4 Pew Research surveys run May-Sept 2018

Outcome: 3-way support for generic Dem, Rep, or Other/Not voting

Simulation: Fix probabilities of missingness and matching as a function of covariates. Sample 2,000 "respondents", then sample subset to match to the file (prop to generated probabilities).



Other approaches

Other approaches

Impute X_{vf}^{U} with conditional VAE:

1. Use matched data to generate realistic observations of X_{vf}^{U} conditional on data observed in the survey $\left(Y^{U}, X_{svy}^{U}, X_{both}^{U}\right)$.
2. Then, fit normal supervised learning method on matched AND imputed data.

Other approaches

Impute X_{vf}^{U} with conditional VAE:

1. Use matched data to generate realistic observations of X_{vf}^{U} conditional on data observed in the survey $\left(Y^{U}, X_{svy}^{U}, X_{both}^{U}\right)$.
2. Then, fit normal supervised learning method on matched AND imputed data.

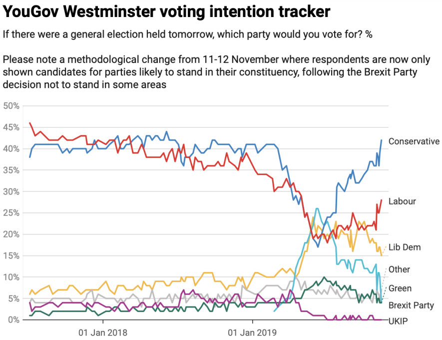
Ecological features: Emulate hierarchical stucture of MRP with embeddings of distributional features

$$Y_i^j = \alpha f(X_i^j) + (1-\alpha) f(\mu_{\mathrm{P}_j})$$

where α controls weight of population v. individual-level features

There's (always) an election!

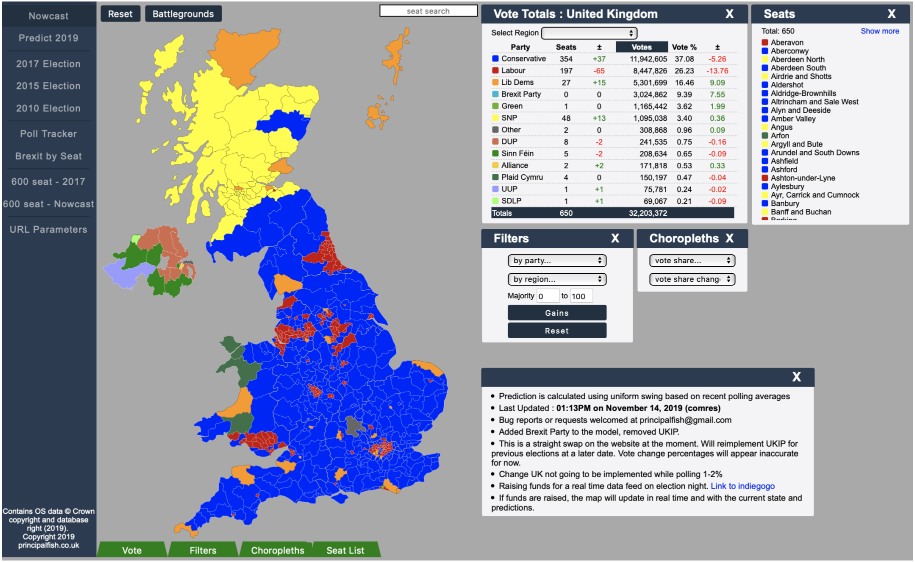
T-27 Days



Some interesting challenges:

- Tactical voting
- Lots of fluidity / some party re-alignment due to Brexit
- Even more privacy restrictions in the UK...basically **no** matched data
- Less advanced analytics infrastructure, very little vote history data

UK Election Maps - Nowcast - Conservative Majority: 58



http://principalfish.co.uk/electionmaps/?map=prediction