Supervised Learning for election forensics with Multi-Agent simulated training data

Fabricio Vasselai
Political Sience & Scientific Computing (vasselai@umich.edu)

Motivation

Generate synthetic training data (STD) for Supervised Machine Learning detection of election fraud, that:

- do not need calibration with (or assumptions of cleanliness in) countries' past elections (see Cantú, 2011; Zhang et.al., 2019)
- minimize the risk of regular strategic behaviour classified as fraud
- allow natural generation of distributions of classifications

MAS synthetic training data

A subfield of Artificial Inteligence, MAS are computer simulations of interacting agents. My MAS of elections:

- implements game-theoretical models in Myerson and Weber (1993), Cox (1994) and Bouton (2013) as iterative discrete-time simulations.
- extends those to include strategic abstention - with ideas from Palfrey and Rosenthal (1985) and Demichelis and Dhillon (2010) - and the presence of sincere voters
- generalizes Myerson's (1998) two-candidate Poisson pivotal probabilities to multi-candidates, with multi-way ties, and to SNTV and runoff

Overview of the SMD case (for simplicity)

Notation 1. Let i be the focal elector, $\mathcal{J}$ be the set of candidates, $j \in \mathcal{J}$ being a candidate who i is considering to vote for. Vector $\mathbf{u}^i \in \mathbb{Q}_{[0,1]}^{|\mathcal{J}|}$ holds individual utilities that i gets in case each candidate wins. π^i is i's current candidate choice; c^i and q^i are i's cost and current probability of voting.

Definition 1. [candidate choosing] $\pi^i = \operatorname{argmax}_{j \in \mathcal{J}'}(E^i_j - E^i_\varnothing)$, where E^i_j and $E^i_\varnothing$ are expected rewards of voting for j or abstaining (i.e.$\varnothing$), and $\mathcal{J}' = \mathcal{J} \setminus \operatorname{argmin}_{h \in \mathcal{J}}(\mathbf{u}^i_h)$.

Definition 2. [updating vote prob.] $q^i = q^i + \operatorname{sign}(\mathbf{u}^i_\pi - c^i) \cdot \epsilon$, where $\epsilon \in \mathbb{Q}_{[0,1]}$ is any small value (as a learning rate; see Mebane et.al., 2019).

Notation 2. Let $\mathcal{T} \in \bigcup_{r=1}^{|\mathcal{J}|-1} \binom{\mathcal{J}\setminus\{j\}}{r}$ be the set of candidates currently tied for 1st rank and $\mathcal{K} = \mathcal{J} \setminus \{j, \mathcal{T}\}$ the set of remaining trailing candidates

Theorem 1. [Expected reward in plurality voting]

$$E^i_j - E^i_\varnothing = \sum_{\mathcal{T}} \left(\frac{\alpha^i_{j,\mathcal{T}}}{|\mathcal{T}|} + \beta^i_{j,\mathcal{T}} \right) \left(\frac{\sum_{t \in \mathcal{T}} (\mathbf{u}^i_j - \mathbf{u}^i_t)}{|\mathcal{T}| + 1} \right)$$

Notation 3. $\alpha^i_{j,\mathcal{T}}$ and $\beta^i_{j,\mathcal{T}}$ are the probabilities that if i votes for j, she will, respectively, create or brake a tie between j and all $t \in \mathcal{T}$. Letting vector $\mathbf{n}^i \in \mathbb{Q}_{\geq 0}^{|\mathcal{J}|}$ hold i's impressions about candidates' expected votes:

Theorem 2. [General pivotal probabilities in plurality voting]

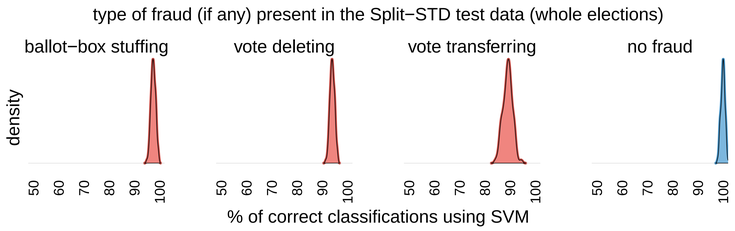
$$\alpha^i_{j,\mathcal{T}} := \Pr\Big(\bigcap_{t \in \mathcal{T}} \mathbf{n}^i_j = \mathbf{n}^i_t \bigcap_{k \in \mathcal{K}} \mathbf{n}^i_j \geq \mathbf{n}^i_k\Big) \quad \beta^i_{j,\mathcal{T}} := \Pr\Big(\bigcap_{t \in \mathcal{T}} \mathbf{n}^i_j = \mathbf{n}^i_t \bigcap_{k \in \mathcal{K}} \mathbf{n}^i_j > \mathbf{n}^i_k\Big)$$

Training schema

- when classifying fraud in whole elections, training was done on the average (across lower levels) number of digits of each candidate's final figures and on the correlation of their vote shares and turnout (Leving et.al., 2009)
- when classifying fraud on lowest levels (e.g.individual constituencies), training was done on the percentages of votes and turnout
- Support Vector Machine (SVM) and Random Forest; Deep Neural Network still on the work. Due to their similarity, only SVM results presented

Split-STD cross-validation

- 10K contests simulated, each with 100 SMD constituencies, 3-6 candidates, 20K-100K electors, 50-90% sincere voters, 20-70% abstention
- repeatedly (for 10K times), applied either no fraud, ballot-box stuffing, vote deletion or vote transferring (25% prob. each)
- each time, split the data into two chunks (20%-80% of original size each) and used one to train and another to test the learner



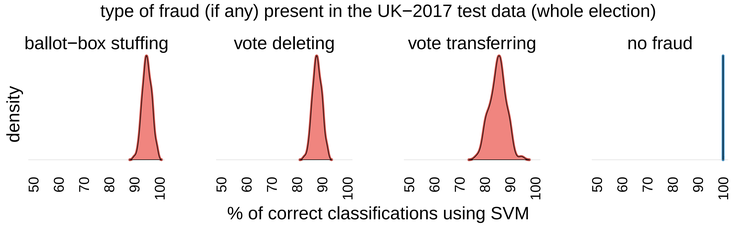
Predicting manually applied fraud on UK-2017

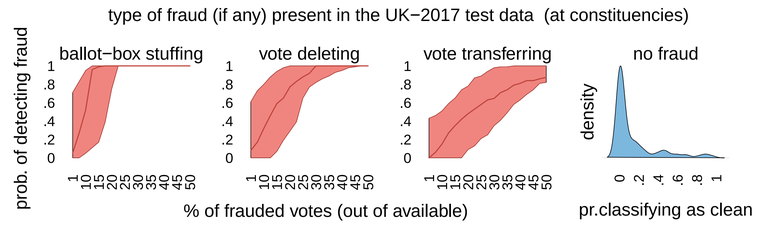
Synthetic Training Data:

- 10K plurality contests simulated via MAS for each of the 650 UK constituencies, following their actual number of candidates and of electors
- actual % turnout of constituencies approximated by setting their MAS doppelganger to have similar % of electors with negative cost of voting
- % of sincere voters per constituency randomly drawn each time, 60%-95%, following actual estimations for the UK (Kiewiet, 2013)
- repeatedly (for 10K times), applied to the STD either no fraud, ballot-box stuffing, vote deletion or vote transferring (25% prob. each)

Test data:

- "manually" manipulated UK-2017 results by constituency, in 50 different ways for each type of fraud (plus 50 instances of clean actual data)



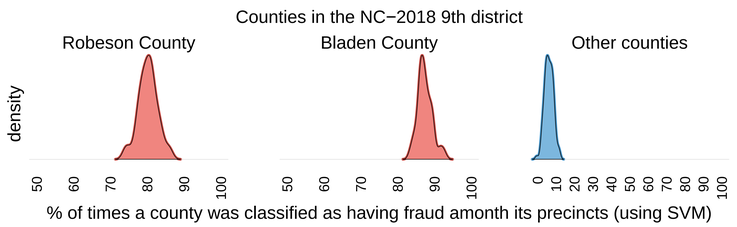


Classifying North Carolina 2018

In 2018, House election in the 9th district of NC was cancelled due to absentee ballots tampered in Counties Robeson and, mostly, Bladen.

Synthetic Training Data:

- 10K plurality contests simulated via MAS for each NC-2018 9th district precincts, following their actual number of electors and candidates
- actual % turnout of precincts are approximated by setting the precinct MAS doppelgangers to have similar share of simulated electors holding a negative cost of voting
- % of sincere voters in each constituency is randomly drawn each time between 80%-95% (there is not much space for strategic behaviour with 2 main candidates and a 3rd with scarce support)
- repeatedly (for 10,000 times), each time applied to the STD either no fraud, ballot-box stuffing or vote deletion (33% prob. each), following description of the actual NC-2018 case



Discussion

- framework particularly good to classify whole elections as having had (or not) ballot-box stuffing or vote deletion
- false positives were rare - less fooled by legit strategic behaviour?
- framework currently significantly worse at classifying lowest levels (constituencies, precincts, etc). Given lack of aggregation in those cases, there is not much more than votes and turnout to train on.
  - idea is to now try keeping only simulated cases whose figures, after fraud is potentially applied, are similar to real
  - weighting STD cases by similarity to real data also showed promise
- testing framework with Mexican elections is underway, since for them, official decisions about frauds are available (i.e.ground truth data).
- the case of Argentina 1932 is also being considered, but besides data being too highly aggregated, the rather particular electoral system requires some tweaks to be simulated through my MAS with SNTV