Bias in signature verification

March 8, 2021

1 Intro

US elections 2020, in mail voting happened in a much larger scale than any previous years. As reported by (1) some 65.6 million mail-in ballots has been cast - double the number in 2016 and more than any other election year. Given that signatures usually reveal the name a person, they also reveal a lot of information about a person's sensitive attributes such as race and gender.

1.1 How does vote validation happen this year?

There has been lots of discrepancy in policies for signature verification among different states. Furthermore, the sates have not been very transparent about their signature verification policy. Based on our research, for each voter's ballot envelop, the signature on the envelop is compared by the signatures (one or many) that exist in the database for the voter. For example, this could be the signature on file from the drivers license. This step is usually performed by a human, whom in most cases might not have extensive training in signature comparison. As reported in (2), about 30 of the top 100 counties use the signature verification software Parascript (link). Parascript declined to identify the users. officials and voting rights advocates mention that the software could reduce the inconsistency and bias inherent when humans decide whether the signature on a ballot envelope matches what governments have on file. Officials and voting rights advocates mention that the software could reduce the inconsistency and bias inherent when humans decide whether the signature on a ballot envelope matches what governments have on file.

Once a ballot is rejected at the first round, then the ballot goes to the next level of human evaluation. This is either the final level of human evaluation or the 2nd out of 3rd. If it's the 2nd out of the 3rd round, then there is only one human evaluator that checks the signature. If it's a "match" (although (3) recommends not using the word match as it brings confusion), then the ballots will be counted. If it's not a match, then the ballot is sent to the 3rd level of verification which is comprised of three humans, one from each of the two dominant parties and one neutral.

The goal of this study is to rigorously measure the effect of the human evaluators' personal biases on the acceptance/rejection of the mail ballots.

2 Human study

Our goal is to design a human study to measure the effect of race on the signature verification accuracy.

2.1 Study outline

- We collect (name, last name) combinations that are highly correlated with certain race or demographic groups. So far we have names that are highly correlated with being African American or White (4)

- We randomly choose a name and show it to the mechanical turker and ask them to write the (name, lastname) as a signature. This step could be implemented in three different ways:
    - A Generate signatures signed on paper from turkers. We can give a set of names to each turker (with repetitions) and ask them to sign it on a piece of paper and upload a photo. Like: Samira Samadi, Bhuvesh Kumar, Stefano Balietti, Samira Samadi, Can Çelebi, Bhuvesh Kumar, Stefano Balietti, Can Çelebi. The repetition allows us to have two signatures of the same name, and we repeat the names with a bit of gap in between so increase the chance of having difference between two signatures of the same name.
    - B Same as before, but instead of signing on a piece of paper and uploading a photo, signing using the trackpad/mouse.
    - C One set of signatures is generated by turkers, using mouse or trackpad. And second set of signatures is generated by AI. For the set generated by AI, we should have multiple AI signature generators so that the user does not see a pattern that reveals that these signatures are AI generated.
- We select a new set of MTurkers. Given the dataset of signature pairs (or single signatures), we randomly show a pair (or a single signature) to the MTurker and ask them if
    - A Show them two signatures and ask "does these two signatures belong to the same person?"
    - B Show them two signatures and ask "does these two signatures belong to the same person?"
    - C Show them one signature and ask "Is this signature generated by a human or is it AI forged?"

2.2 Signature verification training

Found information on guideline, Colorado training document, Oregan training workshop, and LATIMES

3 Information

- Some terms: postal voting or mail-in voting or vote by mail, poll watcher, ballot
- processing of absentee ballots: In many states, processing of absentee ballots can begin before they are actually counted. "Processing" means different things in different states, but typically the first step is comparing the affidavit signature on the outside of the return envelope against the voter's signature on record to ensure a match, or otherwise verify the voter's identity. See the section titled, "Processing, Verifying, and Counting Absentee Ballots." (link)
- Absentee voting: https://www.ncsl.org/research/elections-and-campaigns/vopp-table-14-how-states-verify-voted-absentee.aspx

3.1 Issues with in-mail voting this year

- States have wide latitude over determining election rules, including setting deadlines for a postal vote to qualify. Pennsylvania will only include those received by 20:00 local time on Election Day, while California accepts votes as long as they are postmarked by the date, even if they arrive weeks later. That's why counting in the huge West Coast state always takes a long time.
- Postal ballot counting takes longer because each vote must have a signature that is matched with a separate autograph on a registration card. full list of when the votes will be counted? //www.ncsl.org/research/elections-and-campaigns/vopp-table-16-when-absentee-mail-ballot-processing-and-counting-can-begin.aspx
- Each state can determine the time that the processing of the absentee ballots can start. This varies a lot state to state. For example in some states it can start upon receipt (Georgia, Hawaii, etc) in California Signature verification for all-mail jurisdictions can begin 29 days before Election Day or At 1 p.m. on Election Day, unless a different time, no earlier than two hours after the opening of the polls, is posted and announced at New Hampshire.

3.2 Georgia

- Link: Wikipedia article about 2016 elections (16 Electoral votes)
- Republican: 50.44%
- 45.35%
- 4.21% others (Libertarian, Independent)
- Link: More than 1.6 million voters have requested for mail-in ballot
- Link: GA voting statistics Source of the data state gov website
- Link: Article on how mail in ballots are counted in GA

From sos.ga.gov: Processing Voted Absentee Ballots: Once the information on the oath envelope is verified, the registrar will compare your absentee ballot oath envelope to your voter registration card to verify your signature, as well as compare your signature on the ballot envelope with your signature on the absentee application. On Election Day, the ballot and the envelope are separated to ensure confidentiality of the ballot. This process, and the processing and counting of absentee ballots shall be open to the public. If the absentee ballot is challenged, and that challenge is upheld, the ballot remains in the envelope, is not counted, and the reason is indicated on the envelope.

References

[1] "Us election 2020: When will we get a result and could it be contested?" 2020, bbc.com/news/election-us-2020-54096399.

[2] P. Dave and A. Sullivan, "Factbox: U.s. counties using automated signature verification software," *Reuters*, https://www.reuters.com/article/us-usa-election-ballot-signatures-softwa/factbox-u-s-counties-using-automated-signature-verification-software-idUSKCN26F1U4.

[3] R. Macías, P. Berry, J. Friedmann, J. A. Gomez, C. Fletes, B. Stonesifer, J. Marks, F. Nisen, P. R. Spencer, D. Rose, and K. Alexander, "Re: Proposed regulations," 2020, https://www.aclunc.org/sites/default/files/Final_Regulations_Letter_8-5-20.pdf.

[4] M. Bertrand and S. Mullainathan, "Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination," *American economic review*, vol. 94, no. 4, pp. 991–1013, 2004.