Data science’s impact on election strategies

David Suffolk
Bellevue University
Bellevue, NE 68005
dsuffolk@my365.bellevue.edu

Abstract
In the political campaign landscape, data has increased from broad samples of voter data to voter mapping so specific it can predict voter behavior at the residential level. The sources of data include everything from voter registration to choices of entertainment. The increase of big data has impacted the strategies of political campaigns in advertising targeting to policy building and event planning.

Overall, the current state of data science in political campaigning has seen pros and cons to changes in voter engagement and public policy. If a voter is not in a particular voting block, they may be ignored by their potential representatives. Public policies that are important but not engaging to key voters may not be advocated for by the representative. Alternatively, data has allowed representatives to engage more specifically with the electorate and create policies for issues that they may not have thought of as vital to their constituents.

Author Keywords
Data science, election, political science, public policy data, voter data

ACM Classification Keywords
H.1.m. Information Systems Miscellaneous

History of Voter Registration in the United States

Voter Registration has always been an essential data point for political strategists.

Early 19th Century: Voter registration launched due to concerns from the states about non-citizens voting.

1870-First World War: Several developments in voter registration to resolve conflicts between disenfranchised voters and polling places. The result was an increase in voter participation among the working class, the poor, and immigrants.

1965: Voting Rights Act is passed. Expanded voting and allowed government to investigate discriminatory practices in the voter registration process.

1993: National Voter Registration Act. Requires state governments to offer voter registration to those applying for drivers licenses and public assistance.

Introduction

Data has been an important part of election strategies for a long time. However, the size of the data available has increased. There are potential opportunities for this to advance public policy but also a potential danger in electorates not being represented.

Why is this Data Science?

Political Data Scientists must find the most efficient and effective ways to organize a campaign.[1] These discoveries will target where advertisements are placed, which voters are targeted, policies discussed on the campaign trail, and how to reach donors.[1] The first implementation of pre-election polling with scientific sampling was in 1936 and candidates have used some sort of political polling since then to build a campaign strategy.[2] However, these polls only reveal the intentions of a voter and do not explore the reasons why.[2] Having information about the values of voters can help predict why a voter is stating their intentions the way they do.[2] Campaigns need to know this information so that they can know which voters to target and what messages should be delivered to these voters and their corresponding values.[3]

Campaigns must select data that helps them form an impression of the electorate and their values.[3] The data can come from various sources including public information from the government or from outside entities.[3] However the data is selected and viewed, the available data informs a campaign's decision.[3] The introduction of big data changes the availability of resources for campaigns to use. As a result, new data will reflect a change in politics, which will lead scientists to study political science when they may not have previously, which subsequently lead to new ethical questions and models that deal with these questions.[5]

What is the deliverable for this project?

Since there is more data to be reviewed, there are two concerns that arise about how this new wealth of data is being used by campaigns: voters being ignored because the data segments them from a campaign's target and the quality and use of public policy proposals.

Voter registration records enable campaigns to not only determine who is likely to support them but also who is unlikely to support them which can result in a candidate only focusing on 51% of the constituency.[3] The micro-targeting can subsequently lead to voters not being exposed to alternative political viewpoints.[3]

When it comes to the understanding of public policy, the majority of the electorate are ill-equipped to comprehend the proposals that are being made to them.[4] Additional factors in voter awareness (including candidates with incumbency and campaign spending on advertising) can lead to voters supporting policies that are not in their own interest.[4] In one election that was studied, candidates spoke about policy proposals in short statements on social media and showed little engagement with the follow-up on the proposal.[6]

Proximity Voting Model

One model in understanding voter behavior is the Proximity Voting Model. This model initially takes a voter's values and views and calculates their proximity to one of two candidates. This model was used in reviewing the 2006 election results of the United States House race.

Proximity Rule: $(v_{ij} - D_j)^2 - (v_{ij} - R_j)^2$

The voter's (v) ideals (i) and district (j) are calculated against the Democratic (D) candidate and the Republican (R) candidate. If the Democratic candidate is closer to the voter, the expression is negative. If the Republican candidate is closer to the voter, the expression is positive.

The limitation of the Proximity Voting Model is that voters lack the information to vote along with the Proximity Rule. However, the results showed that informed and uninformed voters were able to vote along the lines of candidates within their Proximity Score. The theory behind this is that voters use party identification as a voting guideline and this typically lines up with the Proximity Rule. Finding a voter's party affiliation is accessible through voter registration records and has been for a while so this has not broken much ground with the evolution of data science.

Perceived Voter Model

The Perceived Voter Model is from the perspective of the campaign and it uses information on a voter to form a campaigns' perception of that voter's likely partisan support, issue support, turnout likelihood and other factors. The campaign then creates what is their ideal perceived voter and targets groups that exhibit those characteristics. The driving idea behind the Perceived Voter Model is getting voters that are like-minded with the candidate to get out and vote (not change their mind about an issue).

Part of any campaign strategy is getting voters to cast their ballots but the Perceived Model begins to present problems and questions about whether data science is impacting representation and engagement between candidates and the electorate. The model is based on assumptions of how a person will vote and enables a campaign to only target those groups. If the campaign has enough data, they may not need to target the voters that they assume will not vote for them.

Public Policy and the Perceived Voter Model

Campaigns can use marketing to promote certain public policy proposals and issues that are important to a candidate. Campaigns have access to such data as voter's party, age, gender, race, and neighborhood characteristics. They also have access to state license records (such as hunting or fishing licenses). This data helps a campaign build an image of a perceived voter and make an assumption of what is important to that voter. Campaigns can then create marketing targeted at the neighborhood level about policy issues that they perceive is important to that neighborhood. There is a risk of false-positives with these set of assumptions.

As discussed in the Proximity Voting Model, most voters are not well-informed about a candidate's positions on a particular public policy or issue. Instead, they largely vote on the assumption that a candidate aligns with a party's position on a policy. The Perceived Voter Model is a more recent approach in using the data available to campaigns but the strategies still have flaws and could end up campaigning to voters they did not intend to target.

Redistricting

In the United States, redistricting is completed every ten years after the submission of the census. New districts determine where local representation resides throughout the country. When a district is more competitive, the campaign can be more stressful on both the competitor and voters.[24] A competitive race can also drive voter engagement because of increased media coverage and higher stakes in the result.[24] There is a correlation between access to increased voter data and a decrease in competitiveness of districts following a redistricting action. The average House District was 1.2-1.4 percent less competitive in 2010 when compared to 1990.[24] While the correlation does not necessarily mean increased data has led to decreased competitiveness, a safer election could now be possible for the representatives that are drawing those district lines. While a campaign may not be fully focused on where a district will be drawn in the next decade or two, the decreased competitiveness adds context to how data science can impact campaigns of the future. If a district can be drawn with little to no competition, there are important questions to be asked about how representatives will engage with their electorate and promote public policy positions of their constituents. If there is no competition, there is evidence that suggests lower voter engagement and media coverage.[24]

Conclusion

The use of models on big data in the Political Science arena are still new and being tested. There are signs of campaigns using the data to take shortcuts on engaging with new voters.[3] While this is appealing because it speaks to the efficiency and effectiveness that campaigns strive for, it is still relatively untested as a campaign strategy.[3] When it comes to engagement with public policy proposals, the messaging has evolved along with social media where campaigns can place a brief synopsis and have little follow-up.[6] However, it is difficult to argue that voter understanding of policies has declined.

Acknowledgements

I would like to thank my professors at Bellevue University for teaching me about Data Science and for providing feedback and guidance on this project.

References

1. Therriault, A. (2016). Finding a place in political data science. *PS: Political Science & Politics, 49*(3), 531-534
2. Cochran, J., Curry, D., Radhakrishnan, R., & Pinnell, J. (2014). Political engineering: optimizing a U.S. Presidential candidate’s platform. *Annals of Operations Research, 215*(1), 63-87.
3. Hersh, E. D. (2015). Hacking The Electorate: How Campaigns Perceive Voters. New York City: Cambridge University Press.
4. Joesten, D. A., & Stone, W. J. (2014). Reassessing proximity voting: expertise, party, and choice in congressional elections. *The Journal of Politics, 76*(3), 740-753.
5. Margetts, H. (2017). The data science of politics. *Political Studies Review, 15*(2), 201-209.
6. Laaksonen, S., Nelimarkka, M., Tuokko, M., Marttila, M., Kekkonen, A., & Villi, M. (2017). Working the fields of big data: Using big-data-augmented online ethnography to study candidate–candidate interaction at election time. *Journal of Information Technology & Politics*, 14(2), 110-131.
7. Dalton, C. M., & Stallman, T. (2018). Counter-mapping data science. *Canadian Geographer, 62*(1), 93-101.
8. Davidson, N. M. (2017). Affordable housing law and policy in an era of big data. *Fordham Urban Law Journal, 44*(2), 277-300.
9. Kumar, K. A., & Natarajan, S. (2016). Role of social media in political campaigning and its evaluation

methodology: a review. *Global Management Review, 10*(3), 1-12.

10. Linzer, D. A. (2014). The future of election forecasting: more data, better technology. *PS: Political Science & Politics, 47*(2), 326-328.
11. Van Der Straeten, K., Sauger, N., Laslier, J., & Blais, A. (2013). Sorting out mechanical and psychological effects in candidate elections: an appraisal with experimental data. *British Journal of Political Science, 43*(4), 937-944.
12. Xie, Z., Liu, G., Wu, J., & Tan, Y. (2018). Big data would not lie: prediction of the 2016 Taiwan election via online heterogeneous information. *EPJ Data Science, 7*(1), 1-16.
13. Hassanpour, Navid. (2013). Tracking the semantics of politics: a case for online data research in political science. PS: Political Science and Politics, 46(2), 299-306.
14. Schoonvelde, M., Schumacher, G., & Bakker, B. (2019). Friends with text as data benefits: assessing and extending the use of automated text analysis in political science and political psychology. Journal of Social and Political Psychology, 7(1), 124-143.
15. Schatz, E., & Maltseva, E. (2012). Assumed to be universal: the leap from data to knowledge in the "American Political Science Review". Polity, 44(3), 446-472.
16. Jungherr, A., & Theocharis, Y. (2017). The empiricist's challenge: asking meaningful questions in political science in the age of big data. Journal of Information Technology & Politics, 14(2), 97-109.
17. King, G., Honaker, J., Joseph, A., & Scheve, K. (2001). Analyzing incomplete political science data: an alternative algorithm for multiple imputation. The American Political Science Review, 95(1), 46-69.
18. Gesing, F. (2018). Transnational municipal climate networks and the politics of standardisation: the contested role of climate data in the new global covenant of mayors for climate and energy. Politics and Governance, 6(3), 126-135.
19. Isaac, J. C. (2014). Restructuring the social sciences? A reflection from the editor of "Perspectives on Politics". *PS: Political Science and Politics*, 47(2), 279-283.
20. Earl, J. (2018). The promise and pitfalls of big data and computational studies of politics. *Partecipazione e Conflitto*, 11(2), 484-496.
21. Giest, S., & Ng, R. (2018). Big data applications in governance and policy. *Politics and Governance*, 6(4), 1-4.
22. Crasnow, S. (2012). The role of case study research in political science - evidence for causal claims. *Philosophy of Science*, 79(5), 655-666.
23. Berman, F., Rutenbar, R., Hailpern, B., Christensen, H., Davidson, S., Estrin, D., Franklin, M., Martonosi, M., Raghavan, P., Stodden, V., & Szalay, A. S. (2018). Realizing the potential of data science. *Communications of the ACM*, 61(4), 67-72.
24. Seabrook, N. R. (2017). Drawing the Lines: Constraints on Partisan Gerrymandering in U.S. Politics. New York City: Cornell University Press.