Agent-based Simulation of District-based Elections with Heterogeneous Populations

Extended Abstract

Adway Mitra

Indian Institute of Technology Kharagpur, India

adway@cai.iitkgp.ac.in

ABSTRACT

In district-based elections, voters cast votes in their respective districts. In each district, the party with maximum votes wins the corresponding "seat" in the governing body. The election result is based on the number of seats won by different parties. In this system, locations of voters across the districts may severely affect the election result even if the total number of votes obtained by different parties remains unchanged. A less popular party may win more seats if their supporters are suitably distributed spatially. This happens due to various regional and social influences on individual voters which modulate their voting choice, especially in heterogeneous societies. In this paper, we explore agent-based models for district-based elections, where we consider each voter as an agent, and try to represent their social and geographical attributes and political inclinations using probability distributions. We propose several models which aim to represent one or more of these aspects. These models can be used to simulate election results by Monte Carlo sampling. The models allow us to explore the possible outcomes of an election, and can be calibrated to actual election results for suitable values of parameters obtained by Approximate Bayesian Computation. Our model can reproduce results of elections in India and USA, and also simulate counterfactual scenarios.

KEYWORDS

Agent-based Modeling, Computational Social Choice

1 OVERVIEW

The aim of this work is to develop agent-based models to simulate voter behavior in a district-based election, that is used in many democratic systems. Significant research has gone into understanding voter behavior, particularly in heterogeneous societies where political choices are related to social identity [3, 4, 6, 7, 10]. The result of elections, i.e. seats won by different parties is strongly dependent on the spatial distribution of voters and districts [2, 5, 8, 9, 11–13, 17]. Agent-based Modeling is an approach of simulation of complex systems, which has been used extensively in epidemiology [1], ecology [14] and economics [18]. This is one of the first works to

build a detailed agent-based model for voter behavior. Here, each voter is treated as an agent with certain attributes, and their votes are a function of their interactions, represented by probabilistic models. The goal of the simulation is to explore possible outcomes of an election under such circumstances.

Let the total number of districts be S, each of which has a seat in the governing body. Voter i belongs to community C_i and votes in district Z_i. The districts have $\{n_1, n_2 \ldots, n_S\}$ voters, with $n_1 + n_2 + \cdots + n_S = N$. Now, there are K political parties with $\{v_1, v_2, \ldots, v_K\}$ supporters, such that $v_1 + v_2 + \cdots + v_K = N$. The vote proportion of these parties can be considered as a K-dimensional discrete distribution, θ. In an election, denote the number of votes for the parties in district s by $\{V_{s1}, V_{s2}, \ldots, V_{sK}\}$. The winner W_s is the party with the highest number of votes in that district. The number of seats M_k won by party k is the number of districts where it is the winner. In the proposed models, some or all of Z, V, W, C and M are considered as random variables.

2 AGENT-BASED MODELS FOR VOTERS

We discuss several agent-based models of voter behavior in the above setting. Each model captures some aspects of voter behavior.

District-wise Polarization Model (DPM): Here we consider the effect of local polarization, where in each district the voters choose a party based on local popularity. If n_{sk} voters in district s already supported party k, a new voter in that district will either choose k with probability $\gamma_s n_{sk}$, or choose according to the overall popularity θ. Here γ_s is the *polarization parameter* for district s. This model represents a trade-off between the popularities of local candidates and the top leadership of parties. This model is based on the famous Chinese Restaurant Process (CRP) [16].

Party-wise Concentration Model (PCM): The effect of this model is to create local concentrations of support in favour of different parties, which helps them to be effective in district-based elections. It is inspired by the fact that support to political parties is often based on social identities, and people often choose residential areas based on social identities. For this model, we once again use Chinese Restaurant Process. But this time we make the process two-step: each person i is first assigned to a party X_i, then (s)he is assigned a district Z_i with a probability proportional to number of supporters of X_i in that district. The proportionality constants $\{\eta_1, \ldots, \eta_K\}$ are the *concentration parameters* for the parties. High value of the parameter η_k encourages voters of party k to concentrate in a few districts, instead of spreading out uniformly.

Social Identity Model: In this model, we explicitly consider the community-based identities of the voters. There are C social communities, and η_c denotes the proportion of people from community c. η is sampled from a Stick-breaking prior. To every person i, we

assign their community as $C_i \sim Categorical(\eta)$. The people from the same community tend to stay together in the same district. Each person i is assigned to district Z_i by following a Chinese Restaurant Process [16] with parameter α. Person i, resides in district s with probability proportional to $\alpha n_s(C_i) = \alpha \sum_{j=1}^{i-1} \mathbb{I}(C_j = C_i)\mathbb{I}(Z_j = s)$ (i.e. number of people from same community as i already residing in district s), or resides in any district chosen uniformly at random with probability proportional to $(1 - \alpha)$. Each community is associated with a prior over the political preferences of its members. For community c and party k, we assign $\phi_{ck} \in \{-1, 0, 1\}$, indicating if the relation between them is bad (-1), neutral (0) or good (1) according to some process or distribution f. Also, a variance σ_k is associated with each party which may be drawn from a Gamma distribution. Finally, for each voter i, their valuation of party k is denoted by $\lambda_{ik} \sim \mathcal{N}(\phi_{ck}, \sigma_k)$ where $c = C_i$. In an election each voter casts their votes on the basis of these valuations. We also take into account *Local Influence*, as the i-th voter can combine their own valuations with the mean valuations of other voters in the same district using a weighing factor κ_i that follows Beta prior.

3 EXPLORING POSSIBLE OUTCOMES

Next, we explore the simulations by these models in a 3-party system with $S = 100$ districts and $N = 1000000$ voters, and considering a few values of the popularity proportion θ. The aim is to see how the parameter values can impact the election results by altering the distribution of voters, even if the popularity proportion of the different parties is fixed. For each setting, the number of seats won by the different parties is noted by averaging across 100 runs.

In case of DPM, it is found that low values of concentration causes almost all seats to go to the most popular party, while high concentration causes the seat share to approach θ. With moderately high values of concentration, and when popularity proportions of the parties are comparable, we find potentially tight results. But the more interesting situations arise in case of PCM, where different parties can have different concentration values, resulting in situations where a less popular party can outperform a more popular one in some cases. This is a characteristic of multi-party elections in different countries such as USA and India. When all three parties have low concentration, the most popular party tends to win almost all the seats, and when all three parties have high concentration the seat share is similar to θ. However, when the popularity proportions are comparable but concentrations are different, the results are most fascinating (Table 1).

In case of the Social Identity model, we consider four scenarios - two involving 3 communities, and two more involving 5 communities, with varying sizes indicated by η. In each case, Scenario 1 (polarized) involves a party that is favored ($\phi = 1$) by the largest communities and opposed ($\phi = -1$) by the smaller ones, one party that is favored by the smaller communities and opposed by the largest ones, and a third party which is neutral ($\phi = 0$) to all communities. The third party has $\sigma = 2$, while the rest have $\sigma = 1$. In Scenario 2 (non-polarized), each party is favored by one or more communities, but not opposed ($\phi = 0$) by the rest. One party again has high $\sigma = 2$, the others have $\sigma = 1$. It is seen that Scenario 1, the centrist/neutral party fails to win any seat with 3 communities, but

θ	$\eta = (0.99, 0.99, 0.50)$			$\eta = (0.99, 0.50, 0.99)$		
	nA	nB	nC	nA	nB	nC
(0.5, 0.4, 0.1)	56	42	2	51	48	1
(0.4, 0.3, 0.3)	43	26	31	42	31	27
(0.4, 0.35, 0.25)	44	35	21	42	41	18
(0.37, 0.33, 0.3)	36	32	32	37	37	26

Table 1: Number of seats won by 3 parties (A,B,C) under different parameter settings of Partywise Concentration Model, for various popularity proportions θ.

	observed			simulated		
Year	M1	M2	M3	M1	M2	M3
2019-1	0.48	0.34	0.18	0.51	0.31	0.18
2019-2	0.45	0.40	0.15	0.43	0.41	0.16
Year	V1	V2	V3	V1	V2	V3
2019-1	114	23	10	113	28	6
2019-2	88	52	7	89	58	0

Table 2: Comparison of observed and simulated results for 2 simultaneous elections in Odisha state, 2019 using Social Identity Model. Above: rounded popular vote shares (M1,M2,M3) of 3 main parties, below: seats won (V1,V2,V3).

can do well with 5 communities. Local influence is found to benefit the parties that support the larger communities and harms the centrist party, particularly when fewer communities are involved.

4 SIMULATION OF OBSERVED RESULTS

It is important to validate the above models to show that they are capable of producing realistic results. For this purpose, we attempt to simulate actual multi-party elections in India and USA, and check if we can reproduce their results. For this purpose, it is necessary to estimate optimal values of the parameters. As parameter estimation techniques like Expectation-Maximization are not applicable, we utilize Approximate Bayesian Computation (ABC). Here, we explore the parameter space by running simulation with a set of parameter values, and accepting them if the produced result is close enough to the actual results. We fit the models to Indian state assembly elections in Delhi-NCR and Odisha, both of which have tripartite contests. We run DPM and PCM by providing them with θ: the vote-share of different parties, and find that the PCM is particularly suitable in reproducing the number of seats won by the parties under optimal parameter settings found by ABC based on a few elections. These settings can also be extrapolated to estimate the results of other elections. We are also able to fit PCM to USA Presidential Elections 2016 and 2020. In all cases, we can explore alternate results of these elections if the geographical distribution of the voters had been different.

In case of Odisha, we ran the SIM by specifying the η and ϕ variables (community-party relations) based on post-poll surveys. It turns out that the popular vote proportions and seat proportions, as simulated by SIM, are reasonably close enough to the actual results, as shown in Table 2. This shows that SIM can simulate realistic results. For more detailed analysis of these models, please see the full version of the paper [15].

REFERENCES

[1] Abe, M., Tanaka, F., Pereira Junior, J., Bogdanova, A., Sakurai, T., and Aranha, C. Using agent-based simulator to assess interventions against covid-19 in a small community generated from map data. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*, pp. 1–8, 2022.

[2] Borodin, A., Lev, O., Shah, N., and Strangway, T. Big city vs. the great outdoors: Voter distribution and how it affects gerrymandering. In *IJCAI*, pp. 98–104, 2018.

[3] Braha, D. and De Aguiar, M. A. Voting contagion: Modeling and analysis of a century of us presidential elections. *PloS one*, 12(5):e0177970, 2017.

[4] Brooks, C., Nieuwbeerta, P., and Manza, J. Cleavage-based voting behavior in cross-national perspective: Evidence from six postwar democracies. *Social Science Research*, 35(1):88–128, 2006.

[5] Chen, J., Rodden, J., et al. Unintentional gerrymandering: Political geography and electoral bias in legislatures. *Quarterly Journal of Political Science*, 8(3):239–269, 2013.

[6] Dawkins, C. J. Measuring the spatial pattern of residential segregation. *Urban Studies*, 41(4):833–851, 2004.

[7] Dawkins, C. J. Space and the measurement of income segregation. *Journal of Regional Science*, 47(2):255–272, 2007.

[8] DeFord, D., Duchin, M., and Solomon, J. A computational approach to measuring vote elasticity and competitiveness. *Statistics and Public Policy*, 7(1):69–86, 2020.

[9] Erdélyi, G., Hemaspaandra, E., and Hemaspaandra, L. A. More natural models of electoral control by partition. In *International Conference on Algorithmic DecisionTheory*, pp. 396–413. Springer, 2015.

[10] Katz, J. N., King, G., and Rosenblatt, E. Theoretical foundations and empirical evaluations of partisan fairness in district-based democracies. *American Political Science Review*, 114(1):164–178, 2020.

[11] Lasisi, R. O. Improved manipulation algorithms for district-based elections. In *The Thirty-First International Flairs Conference*, 2018.

[12] Lev, O. and Lewenberg, Y. "reverse gerrymandering": Manipulation in multi-group decision making. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 2069–2076, 2019.

[13] Lewenberg, Y., Lev, O., and Rosenschein, J. S. Divide and conquer: Using geographic manipulation to win district-based elections. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pp. 624–632, 2017.

[14] McLane, A. J., Semeniuk, C., McDermid, G. J., and Marceau, D. J. The role of agent-based models in wildlife ecology and management. *Ecological modelling*, 222(8):1544–1556, 2011.

[15] Mitra, A. Agent-based simulation of district-based elections. *arXiv preprint arXiv:2205.14400*, 2022.

[16] Pitman, J. Exchangeable and partially exchangeable random partitions. *Probability theory and related fields*, 102(2):145–158, 1995.

[17] Stoica, A.-A., Chakraborty, A., Dey, P., and Gummadi, K. P. Minimizing margin of victory for fair political and educational districting. *arXiv preprint arXiv:1909.05583*, 2019.

[18] Todd, A., Beling, P., Scherer, W., and Yang, S. Y. Agent-based financial markets: A review of the methodology and domain. In *2016 IEEE symposium series on computational intelligence (SSCI)*, pp. 1–5. IEEE, 2016.