Strategic Voting with Incomplete Information

Ulle Endriss University of Amsterdam
Svetlana Obraztsova Hebrew University
Maria Polukarov University of Southampton
Jeff Rosenschein Hebrew University

Abstract: Classical results in social choice theory on the susceptibility of voting rules to strategic manipulation make the assumption that the manipulator has complete information regarding the preferences of the other voters. In reality, however, voters only have incomplete information, which limits their ability to manipulate. We explore how these limitations affect both the manipulability of voting rules and the dynamics of systems in which voters may repeatedly update their own vote in reaction to the moves made by others. We focus on the plurality, veto, k-approval, Borda, Copeland, and maximin voting rules, and consider several types of information that are natural in the context of these rules, namely information on the current front-runner, on the scores obtained by each alternative, and on the majority graph induced by the individual preferences.

Preliminaries

Set of *voters* $N = \{1, \ldots, n\}$ and set of *candidates* C, with $|C| = m$.

True *preferences* $\succ_i$ and declared *ballots* b_i are linear orders, in $\mathcal{L}(C)$.

Resolute *voting rule* $F : \mathcal{L}(C)^n \to C$ to pick a single winner.

To ensure resoluteness, we use *lexicographic tie-breaking*.

Focus on this poster is on *Copeland* and *positional scoring rules*, including in particular plurality, veto, and other k-approval rules.

Safe Manipulation under Uncertainty

Information function π mapping profile $\boldsymbol{b}$ to "information" $\pi(\boldsymbol{b})$, e.g: winner information, score information, or majority graph information.

Given signal $\pi(\boldsymbol{b})$, voter i must consider these partial profiles possible:

$$\mathcal{W}_i^{\pi(\boldsymbol{b})} = \{\boldsymbol{b}'_{-i} \in \mathcal{L}(C)^{n-1} \mid \pi(b_i, \boldsymbol{b}'_{-i}) = \pi(\boldsymbol{b})\}$$

She might manipulate by voting $b_i^\star$ instead of b_i if both:

- $F(b_i^\star, \boldsymbol{b}^\star_{-i}) \succ_i F(b_i, \boldsymbol{b}^\star_{-i})$ *for some* $\boldsymbol{b}^\star_{-i} \in \mathcal{W}_i^{\pi(\boldsymbol{b})}$
- $F(b_i^\star, \boldsymbol{b}'_{-i}) \succcurlyeq_i F(b_i, \boldsymbol{b}'_{-i})$ *for all* $\boldsymbol{b}'_{-i} \in \mathcal{W}_i^{\pi(\boldsymbol{b})}$

Results on Manipulability

The general spirit of Gibbard-Satterthwaite prevails. Still, there are some exceptions (see Reijngoud & Endriss, AAMAS-2012, and this paper):

- Given *winner* information, *veto* is immune to manipulation.
- Given *majority graph* information, the *k-approval* rules with $k \leqslant m-2$ (i.e., all but veto) are immune to manipulation.

Iterative Voting

Iterative voting with voting rule F under information function π:

- *initialise:* all voters vote truthfully $[b_i^0 := \succ_i]$
- then *repeat:* some voter i manipulates $[\boldsymbol{b}^{k+1} := (b_i^\star, \boldsymbol{b}^k_{-i})]$

Will this process *converge* to a *stable outcome*?

Related work: for full-information case, only rules known to converge are plurality and veto (under best-response dynamics).

Convergence Results

Theorem 1. *When voters are given only winner information, iterative Copeland voting always converges to a stable outcome.*

Proof sketch: Under Copeland, no voter will ever want to promote her *least favourite candidate* $c^\perp$ from the *bottom spot* (as $c^\perp$ might win).

Define the *refined Copeland score* of candidate c in profile $\boldsymbol{b}$, taking into account c's ranking in the tie-breaking order:

$$\textit{score}(c, \boldsymbol{b}) + \frac{m - \textit{index}(c)}{m+1}$$

Refined score of current winner $c^\star$ *cannot decrease*, because that of manipulator's $c^\perp$ (who *could* be current runner-up) cannot decrease.

Theorem now follows from standard *potential argument*. ✓

Theorem 2. *When voters are given only winner information, iterative PSR voting always converges to a stable outcome—if voters are "conservative" and only make minimal updates (in terms of Kendall tau distance).*

Proof sketch: Minimal updates rule out the promotion of someone's least favourite candidate. Rest of the proof is the same. ✓