Doctopic: Algorithms, automata, complexity and games ARTICLE IN PRESS TCS:11528

Theoretical Computer Science ••• (••••) •••–•••

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Manipulative elicitation – A new attack on elections with incomplete preferences

Palash Dey

Indian Institute of Technology, Kharagpur, India

A R T I C L E I N F O

Article history:
Received 14 November 2017
Accepted 29 March 2018
Available online xxxx
Communicated by X. Deng

Keywords:
Computational social choice
Voting
Manipulation
Elicitation
Incomplete election
Algorithm
NP-completeness

A B S T R A C T

Lu and Boutilier [22] proposed a novel approach based on "minimax regret" to use classical score based voting rules in the setting where preferences can be any partial (instead of complete) orders over the set of alternatives. We show here that such an approach is vulnerable to a new kind of manipulation which was not present in the classical (where preferences are complete orders) world of voting. We call this attack "manipulative elicitation." More specifically, it may be possible to (partially) elicit the preferences of the agents in a way that makes some distinguished alternative win the election who may not be a winner if we elicit every preference completely. More alarmingly, we show that the related computational task is polynomial time solvable for a large class of voting rules which includes all scoring rules, maximin, Copeland$^{\alpha}$ for every $\alpha \in [0,1]$, simplified Bucklin voting rules, etc. We then show that introducing a parameter per pair of alternatives which specifies the minimum number of partial preferences where this pair of alternatives must be comparable makes the related computational task of manipulative elicitation NP-complete for all common voting rules including a class of scoring rules which includes the plurality, k-approval, k-veto, veto, and Borda voting rules, maximin, Copeland$^{\alpha}$ for every $\alpha \in [0,1]$, and simplified Bucklin voting rules. Hence, in this work, we discover a fundamental vulnerability in using minimax regret based approach in partial preferential setting and propose a novel way to tackle it.

1. Introduction

Aggregating preferences of a set of agents over a set of alternatives is a fundamental problem in voting theory which has been used in many applications in AI for making various decisions. Prominent examples of such applications include collaborative filtering [26], similarity search [18], winner determination in sports competitions [2], etc. [25]. In a typical scenario of voting, we have a set of alternatives, a tuple of "preferences", called a profile, over the set of alternatives, and a voting rule which chooses a set of alternatives as winners based on the profile. Classically, preferences are often modeled as complete orders over the set of alternatives. However, in typical applications of voting in AI, collaborative filtering for example, the number of alternatives is huge and we have only partial orders over the set of alternatives as preferences.

There have been many attempts to extend the use of voting theory in settings with incomplete preferences. The approach of Konczak and Lang [21] was to study the possible and necessary winner problems. In these problems, the input is a profile of partial preferences and we want to compute the set of alternatives who wins (under some fixed voting rule) in at least one completion of the profile for the possible winner problem; for the necessary winner problem, we want to compute the

E-mail address: palash@tifr.res.in.

https://doi.org/10.1016/j.tcs.2018.03.028

Please cite this article in press as: P. Dey, Manipulative elicitation – A new attack on elections with incomplete preferences, Theoret. Comput. Sci. (2018), https://doi.org/10.1016/j.tcs.2018.03.028

set of alternatives who wins in every completion of the profile. There have been substantial research effort in the last decade to better understand these two problems [23,28–30,3,5,1,4,24,19,10,9,17,8,7]. One of the main criticisms of this approach is that the definition of a necessary winner is so strong that none of the alternatives may satisfy it whereas the definition of a possible winner is so relaxed that a large number of alternatives may satisfy it. Moreover, the computational problem of finding the set of possible winners is NP-hard for most of the common voting rules (finding the set of necessary winners is also co-NP-hard for some voting rules, ranked pairs for example) [30].

Lu and Boutilier [22] took a completely different approach to handle incomplete preferences and proposed a worst case regret based approach for score based voting rules. These voting rules assign some score to every alternative based on the profile and select the alternatives with the maximum (or minimum) score as winners. Many popular voting rules, for example, scoring rules, maximin, Copeland, etc. are score based voting rules. For score based voting rules, intuitively speaking, the worst case regret, called maximum regret in [22], of declaring an alternative w as a winner is the maximum possible difference between the score of w and the score of a winning alternative in any completion of the input partial profile; the winners of a partial profile are the set of alternatives with the minimum maximum (called minimax) regret. A completion of a partial profile is another profile where every preference is complete and it respects the orderings of the corresponding preference in the partial profile. The minimax regret based approach is not only theoretically robust as argued in [22] but also practically appealing since computing winners is polynomial time solvable for all commonly used voting rules.

1.1. Motivation

Although the minimax regret based approach enjoys many exciting features, it introduces a new (which was not present in the classical setting with complete preferences) kind of attack on the election which we call "manipulative elicitation." That is, it may be possible to partially elicit the preferences in such a way that makes some favorable alternative win the election. For example, let us consider a plurality election $\mathcal{E}$ where an alternative, say w, is the top alternative of one preference and another alternative, say x, is the top alternative of every other preference. In a plurality election, the winners are the set of alternatives who appear as the top alternative in the largest number of preferences. Hence, x is the unique winner in $\mathcal{E}$. Let us now consider a partial profile where, in every partial preference, only w and every other alternative who is preferred less than w in the corresponding preference in $\mathcal{E}$ are comparable. Let us call the resulting partial profile $\mathcal{E}'$. If n is the number of preferences, then the minimax regret plurality score of w in $\mathcal{E}'$ is $(n-1)$ whereas the minimax regret plurality score of every other alternative is n which makes w the unique winner of $\mathcal{E}'$. We call this phenomenon manipulative elicitation. The problem of manipulative elicitation is even more alarming in AI since, in many applications (collaborative filtering for example), the parts of the preferences that will be elicited can often be influenced and controlled in such settings.

1.2. Our contribution

Our main contribution in this paper is the discovery of the manipulative elicitation attack in regret based partial preferential setting. We also show that the corresponding computational problem for manipulative elicitation is polynomial time solvable for every *monotone* voting rule which includes all commonly used score based voting rules [Theorem 3.1 and Corollary 3.1]. Intuitively speaking, we call a score based voting rule monotone if improving the position of some alternative in any (complete) preference can only improve its score; we defer its formal definition till Section 2. To counter the negative result of Theorem 3.1, we introduce a parameter per pair of alternatives which specifies the minimum number of partial preferences where these two alternatives should be comparable. We establish success of our approach by showing that the new constraints make the corresponding computational task of manipulative elicitation NP-complete for a large class of scoring rules [Theorem 4.1] which includes the plurality [Theorem 4.2], veto [Theorem 4.3], k-approval for any k, and Borda voting rules [Corollary 4.1], maximin [Theorem 4.4], Copeland$^\alpha$ for every $\alpha \in [0,1]$ [Theorem 4.5], and simplified Bucklin [Theorem 4.6] voting rules. We remark that there could be various ways to enforce lower bounds on the number of partial preferences where a particular pair of alternatives is comparable. For example, this can be a feature in the applications which would allow users to generate these bounds from some distribution which would in turn overrule the possibility of such manipulation (due to our hardness results).

1.3. Related Literature

A line of work that is partially related to ours is "preference elicitation" where the goal is to elicit preferences of a set of agents by asking them as few number of queries as possible (called query complexity). Indeed such a reduction in query complexity is shown for single peaked [13], singe peaked on trees [12], single crossing [14] profiles and domains with having small top cycle [11]. Another related line of work is about eliciting few random preferences to predict winner [16] and estimate margin of victory [15].

Organization The rest of the paper is organized as follows. In Section 2, we introduce notion and define our computational problems. We then present our polynomial time algorithm in Section 3 and hardness results in Section 4. We finally conclude with future direction of research in Section 5. A preliminary version of this work appeared at AAAI-2018 [6].