Committee Voting with Incomplete Approval Under Constraints

Ido Amit
Advisor: Prof. Benny Kimelfeld

February 20, 2022

Abstract

When we think of an election with an incomplete voting profile, we consider the necessary winners and the possible winners. Necessary winners are candidates who win for every completion of the voting profile, while possible winners are candidates who win for at least one completion. Yet, our interest might go beyond the identity of winners, rather in their properties or whether they meet a required condition. Another aspect of the problem is to estimate the quality of a set of winners, which we'll describe later. This research discusses alternatives to apply constraints over a set of winners.

1 Introduction

One of the study fields in social choices is the determination of winners. Winners could be candidates in an election, the most recommended titles to watch in a streaming application, or the highest-ranking sites in search engines for a given query[1]. A central issue is dealing with partial voting profiles. We denote a necessary winner if he wins in every completion of the voting profile. In The same manner, we denote a possible winner if he wins in any completion [2]. In some scenarios, we might want to choose a set of winners, i.e., a committee. A committee is a subset of candidates, chosen as winners in a multi-winner election setting [5]. This situation raises a few questions, such as, how we should choose its' members, or how we can rank the quality of a committee. For that purpose, we'll define ranking functions (weight function), to measure the quality of a committee [3]. In this context, we'll focus on ABC rules (Approval-Based Committee) [4]. At certain times, life can be more complex, and it brings upon us constraints over a set of winners, i.e., committee members. For example, two members of a committee can not be relatives. Another example could for be every two candidates from the same party, to agree on a specific issue. We notice our first constraint discussed the relation between members in the committee, while the other one discussed the relation between the committee members and some property. In this research, we will describe alternatives to express constraints over committees, such as first-order logic [6], and functional dependencies [7].

2 Preliminaries

2.1 Relational Databases and Conjunctive Queries

A relational schema is a sequence of distinct finite attribute names $(\mathrm{att}_1, \ldots, \mathrm{att}_k)$, each with a domain of legal values $(\mathrm{dom}_1, \ldots, \mathrm{dom}_k)$. A tuple is a sequence of values $(v_1, .., v_k)$, where each $v_i \in \mathrm{dom}_i$. A relation is a pair (s,r) where s is a relational schema, and r is a finite set of tuples over s. A signature is a finite set of relation names, each mapped into a relation schema. A schema is a signature and constraints. Finally, we can define a database. A database is an instantiation of relation for each relational schema over a schema. A query q, is a function that maps each database D, to a relation R. We denote it $q(D)$ as the result relation of evaluation q over D, and each tuple is $q(D)$ referred to as an answer to the question q in D [10]. We'll study conjunctive queries of first-order logic built from atomic formulas, to express constraints over members of the committee, or committee composition.

2.2 Functional Dependencies

A Functional Dependency (FD for short) over a relation schema S is an expression of the form $X \to Y$, where X and Y are sets of attributes of S. We may also write X and Y by simply concatenating the attribute symbols; for example, we may write $AB \to C$ instead of $A \cup B \to C$ for the relation schema R(A, B, C). An FD $X \to Y$ is trivial if $Y \subseteq X$, and otherwise it is nontrivial. A relation r satisfies FD $A \to B$ if for every pair of tuples t_1, t_2 where $t_1[A] = t_2[A]$ then $t_1[B] = t_2[B]$. In the same manner, a relation r satisfies a set of FD, F, if $\forall f \in F : r \models f$ [15].

2.3 Incomplete Databases and Possible Worlds

Given incomplete database D, meaning lack of information, we denote a possible world of D as a legal completion of D into D', which is a complete database [11, 12, 13]. Query answering is a central challenge in the context of incomplete databases, due to its computational hardness, where the target is to find certain answers and possible answers. We denote a certain answer as an answer of $q(D)$ if, for every legal completion D' of D, it's an answer of $q(D')$. In the same manner, we denote a possible answer if it's an answer for some possible world, i.e., for any completion D' of D, it's an answer of $q(D)$ [10]. As described in Benny Kimelfeld's paper [10], if W denotes the set of possible worlds of incomplete database D, then the set of certain answers is $\cap_{D' \in W} q(D')$, while the set of possible answers is $\cup_{D' \in W} q(D')$. From now on we denote necessary winners as NW, while possible winners as PW.

2.4 Voting Profile and Voting Rules

2.4.1 Positional Scoring Rules

Let $C = c_1, \ldots, c_m$ be a set of candidates and $V = v_1, \ldots, v_n$ be a set of voters. We denote $T = T_1, \ldots, T_n$ a complete voting profile, if each T_i is a total order of set C. meaning $T_i = (c_{i_1}, c_{i_2}, \ldots, c_{i_m})$. The order means the preference of voter v_i over candidates, while c_i comes before c_j means voter i prefer candidate c_j at least like c_i. As we already know, a voting profile might be incomplete, and in such a case we discuss the completion of it. A completion of partial profile $P = P_1, \ldots, P_n$ is a complete voting profile $T = T_1, \ldots, T_n$ where each T_i simply extends P_i with no repeated candidates [2].

2.4.2 ABC Rules and Committee

Let us denote $A = A_1, \ldots, A_n$ an approval profile, where each $A_i \subseteq C$, meaning A_i is a subset of candidates that a voter v_i approves. Note that C A_i are the candidates of which voter v_i disapprove. An approval-based committee (ABC) rule takes an approval profile and parameter $k \leq m$, and outputs one or more subsets of k candidates. Each subset is referred to as committee. Such a committee w.r.t rule r is a winning committee [5]. Approval profile is a relation, and as such, it might be incomplete. We denote a partial profile $P = P_1, \ldots, P_n$, where each P_i is composed of three sets: $P_i = Top_i \cup Middle_i \cup Bottom_i$. Top_i are the candidates a voter i approves, $Bottom_i$ are those he doesn't approve of, and $Middle_i$ are undecided. In some models, we also have a partial/complete order between $c_i \in Middle_i$, which we'll discuss later regards Benny Kimelfeld's paper [10]. A legal completion of partial profile P = $(\mathrm{P}_1, \ldots, \mathrm{P}_n)$ is a complete approved profile, A = $(\mathrm{A}_1, \ldots, \mathrm{A}_n)$, where $\mathrm{Top}_i \subseteq A_i$ and $\mathrm{Bottom}_i \cap A_i = \emptyset$ [5].

2.4.3 Necessary and Possible Committees

Let us define the notions necessary committee and a possible committee of size k problems, which we'll denote as $NecCom\langle k\rangle$ and $PosCom\langle k\rangle$. Given committee size k, partial profile P, set of candidates C and a set $W \subseteq C$ s.t. $|W| = k$, W is a necessary/possible committee if W is a winning committee for every/any completion of P to A. Meaning, $NecCom\langle k\rangle$ is a determination problem whether W of size k is a necessary committee, and $PosCom\langle k\rangle$ is the problem whether W is a possible committee.

2.5 F-Dependent Committee

Let F be an FD and $K \subseteq C$ be a committee w.r.t to rule r. Let "Candidates" be the relation of candidates and their attributes, and $K - Cand \subseteq Candidates$ the relation corresponding to committee K. We denote a committee K, F-Dependent committee (DC w.t.r to FD F) if the relation "K-Cand" satisfies F. Let us run through an example with the figure below and the FD F:$Party \rightarrow$ Freedom of information. In this example suppose we do like each party in the committee agree on a specific topic, such as freedom of information, and suppose there are two winning committees: $K_1 = \{Harry, Ron, Darco\}$ and $K_2 = \{Harry, Hermione, Darco\}$. In this example, K_1 is an F-Dependent committee while K_2 is not. As we did before, we'll define the notions F-Dependent $NecCom\langle k\rangle$ and $PosCom\langle k\rangle$ for a necessary and a possible F-Dependent committees.

Candidates Relation

candidate name	Party	Freedom of information
Harry	Gryffindor	Supports
Ron	Gryffindor	Supports
Hermione	Gryffindor	Opposes
Draco	Slytherin	Opposes

Figure 1: Figure of Candidates

3 Related Work

The complexity of Computing necessary and possible winners over incomplete voting profiles and notations are studied in Konczak and Lang's paper, which give rise to the classification of positional scoring rules, particularly pure rules. [2]. Adopting the framework of Lackner and Skowron, the properties of ABC rules, as well as algorithms to compute winner committee and computational hardness of doing so have been thoroughly studied [4]. In recent studies, multi-winner elections with diversity constraints and fairness constraints had been explored, where constraints are over the schema's attributes [8, 9]. In paper [8] - the researchers analyzed the computational complexity of computing winning committees, obtained polynomial-time algorithms (exact and approximate) and NP-hardness results. In paper [9] the computational complexity of fairness constrained multi-winner voting problem has been studied for monotone and submodular score functions. In addition, the researchers present several approximation algorithms and matching hardness of approximation results for various attribute group structures and types of score functions. For last, they present simulations that suggest that adding fairness constraints may not affect the scores significantly when compared to the unconstrained case.

4 Formal Framework

Theorem 1
For every functional dependency F, a committee consisting of one candidate

or less is an F-Dependent committee.

4.1 Positional Scoring Rules Over Voting Profile

Positional scoring rule over m candidates is defined by a vector $x = (x_1, \ldots, x_m)$, where $\forall 1 \leq i \leq m : x_i \geq 0$, $\forall i > j$ $x_i \geq x_j$ and $x_m = 0$. Let $C = c_1, \ldots, c_m$ be a set of candidates, $V = v_1, \ldots, v_n$ be a set of voters, r be a positional scoring rule defined by s a scoring vector and $T = T_1, \ldots, T_n$ a complete voting profile [2]. Note each T_i is a complete order of C. The score $s(T_i, c_j)$ is defined as the score x_j in T_i. When scoring r is applied over a complete profile T w.r.t candidate c, each candidate c receive a score as follow: $\sum_{i=1}^{n} s(T_i, c)$. We denote the set of winners, $W(r, T)$, as those who achieved the maximal score.

4.2 Positional Scoring Rules

4.2.1 Pure Rules

Pure rules are a special case of positional scoring rules. A scoring rule is pure if, for each $m \geq 2$, the scoring vector for m candidates can be obtained from the scoring vector for $m-1$ candidates by inserting one additional score value at any position which maintains the decreasing order [14]. Known examples for pure rules are Veto, Plurality, Borda, k-approval, and more [10, 14].

Theorem 2 (Classification Theorem)

1. *For every partial voting profile P and pure positional scoring rule r, computing NW is possible in polynomial time.*
2. *Given partial voting profile P and pure positional scoring rule r, Computing PW is NP-Complete for every r, except for Plurality and Veto rules.*

4.3 ABC Rules Over Partial Approval Profile

4.3.1 Poset

As a reminder, we study the case of a given set of candidates C, partial profile P, where each P_i is composed of Top_i, $Middle$, and $Bottom$ subsets of candidates and a set of voters V, we do like to compute possible and necessary committees. Poset is the general case where we have some partial order $\succ_{v_i}$ over candidates $c \in Middle_i$. Meaning, for two candidates c, c' if $c \succ_{v_i} c'$ and $c' \in A_i$ then $c \in A_i$ as well. Poset can also be interpreted the following way: each voter v_i select a subset of candidates to approve from $Middle_i$ w.r.t partial order $\succ_{v_i}$, will be denoted as M_i. As a result, $A_i = Top_i \cup M_i$ [5].

4.3.2 3VA (Three-Values Approval

3VA (Three-Valued Approval) is a special case of Poset, where the partial order $\succ_{v_i}$ is empty. Meaning, voter v_i can select every subset $M_i \subseteq Middle_i$ and again, $A_i = Top_i \cup M_i$ [5].

4.3.3 Linear

Linear is also a special case of Poset, where the partial order $\succ_{v_i}$ is linear, meaning $\succ_{v_i} = (c_{i_1} \succ c_{i_2} \succ ... \succ c_{i_j})$ where $|Middle_i| = j$. In 3VA each voter v_i select an index $0 \leq k \leq j$ as threshold such that each c_{i_q} which holds $q < k : c_{i_q} \in M_i$, and again, $A_i = Top_i \cup M_i$ [5].

5 Research Goal

The goal of this research is to express constraints over necessary and possible committee's composition in various paths. In details:

1. At first, we introduced the notion of a committee and scoring rules to choose winning committees.
2. Later, we studied different alternatives to express constraints over a committee's composition. Among them are first-order logic and functional dependencies.
3. Then, we defined new notions, such as the F-Dependent committee, to express constraints using functional dependencies, and described the motivation for doing so.
4. Ultimately, we'll discuss the computational challenge in F-Dependent committees and computing the F-Dependent NC and PC.

5. For future research we plan to study the problem in which for a given partial profile, scoring rule, and constraints, among the alternatives above, we wish to find a minimal winning committee or the largest F-Dependent committee.

6 Future Directions

We plan to explore the complexity of finding a minimal winning committee for a given partial profile, scoring rule, and constraints over the committee's composition. We also plan to explore the complexity of finding maximal winning F-Dependent committee. Later, we wish to study polynomial algorithms to compute the minimal winning committee and the maximal F-Dependent committee, as well as Polynomial approximate algorithms. We're also interested in statistic-driven completion of databases, which will use statistic tools, such as imputation, to achieve a complete voting/approval profile in polynomial complexity. In this direction, we trade complexity with certainty and wish to minimize the loss of certainty.

References

[1] Brandt, Felix, et al., eds. Handbook of computational social choice. Cambridge University Press, 2016.

[2] Konczak, Kathrin, and Jérôme Lang. "Voting procedures with incomplete preferences." Proc. IJCAI-05 Multidisciplinary Workshop on Advances in Preference Handling. Vol. 20. 2005.

[3] Cao, Zhe, et al. "Learning to rank: from pairwise approach to listwise approach." Proceedings of the 24th international conference on Machine learning. 2007.

[4] Lackner, Martin, and Piotr Skowron. "Approval-based committee voting: Axioms, algorithms, and applications." arXiv preprint arXiv:2007.01795 (2020).

[5] Imber, Aviram, et al. "Committee Voting with Incomplete Approvals." arXiv preprint arXiv:2103.14847 (2021).

[6] Abiteboul, Serge, and Victor Vianu. "Computing with first-order logic." Journal of computer and System Sciences 50.2 (1995): 309-335.

[7] Fagin, Ronald. "Functional dependencies in a relational database and propositional logic." IBM Journal of research and development 21.6 (1977): 534-544.

[8] Bredereck, Robert, et al. "Multiwinner elections with diversity constraints." Thirty-Second AAAI Conference on Artificial Intelligence. 2018.

[9] Celis, L. Elisa, Lingxiao Huang, and Nisheeth K. Vishnoi. "Multiwinner voting with fairness constraints." arXiv preprint arXiv:1710.10057 (2017).

[10] Kimelfeld, Benny, Phokion G. Kolaitis, and Julia Stoyanovich. "Computational social choice meets databases." arXiv preprint arXiv:1805.04156 (2018).

[11] Fagin, Ronald, et al. "Data exchange: semantics and query answering." Theoretical Computer Science 336.1 (2005): 89-124.

[12] Imieliński, Tomasz, and Witold Lipski Jr. "The relational model of data and cylindric algebras." Journal of Computer and System Sciences 28.1 (1984): 80-102.

[13] Lenzerini, Maurizio. "Data integration: A theoretical perspective." Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems. 2002.

[14] Baumeister, Dorothea, and Jörg Rothe. "Taking the final step to a full dichotomy of the possible winner problem in pure scoring rules." Information Processing Letters 112.5 (2012): 186-190.

[15] Livshits, Ester, Benny Kimelfeld, and Jef Wijsen. "Counting subset repairs with functional dependencies." Journal of Computer and System Sciences 117 (2021): 154-164.