Rank Aggregation by Dissatisfaction Minimisation in the Unavailable Candidate Model

Extended Abstract

Arnaud Grivet Sébert*
Université Paris-Saclay, CEA, List
F-91120 Palaiseau, France
arnaud.grivetsebert@gmail.com

Nicolas Maudet
LIP6, UMR 7606, Sorbonne Université
Paris, France
nicolas.maudet@lip6.fr

Patrice Perny
LIP6, UMR 7606, Sorbonne Université
Paris, France
patrice.perny@lip6.fr

Paolo Viappiani
LIP6, UMR 7606, CNRS and Sorbonne Université
Paris, France
paolo.viappiani@lip6.fr

ABSTRACT

In this paper, we extend the unavailable candidate model [10] and present two new voting rules based on a finer notion of disagreement, called dissatisfaction, which depends on the ranks of the candidates, considered among all the candidates (*ex ante dissatisfaction* rule) or only among the available candidates (*ex post dissatisfaction* rule). We provide algorithmic results for the two rules and show that apparently very different voting rules such as scoring rules or Kemeny rule can be unified under the same aggregation concept: expectation of dissatisfaction under the availability distribution.

KEYWORDS

Computational Social Choice; Preference Aggregation; Unavailable Candidate Model; Polynomial-Time Approximation Scheme

ACM Reference Format:

Arnaud Grivet Sébert, Nicolas Maudet, Patrice Perny, and Paolo Viappiani. 2021. Rank Aggregation by Dissatisfaction Minimisation in the Unavailable Candidate Model: Extended Abstract. In *Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), Online, May 3–7, 2021*, IFAAMAS, 3 pages.

1 INTRODUCTION

While traditional social choice theory assumes that the set of candidates is well known before voting takes place, in recent years, several approaches have been proposed to address the problem of candidates' unavailability [3–6, 9, 11, 12]. An approach of particular interest is *the unavailable candidate model* proposed by Lu and Boutilier [10] where the optimal rankings are computed by minimisation of the expected number of disagreements over all the possible subsets of available candidates. Lu and Boutilier provide a clear decision-theoretic justification for producing a ranking instead of a single winner: the output ranking serves as a very compact decision policy to select the best available candidate as "winner". However, the binary disagreement used in their model or in [1, 7, 8] assumes that a voter is satisfied only if its most preferred available candidate is elected (as in "plurality" rule) and is fully unsatisfied otherwise. We argue that the voter's satisfaction should vary more smoothly and depend on the rank he/she gives to the candidate declared as winner by the aggregation. Moreover, we observe that there are two opposed ways to measure the satisfaction of the voters, either by considering the ranks of the candidates in the whole preference order of the voter (*ex ante* approach), or the ranks of the candidates within the subset of available candidates (*ex post* approach).

2 BACKGROUND

Given a set $\mathscr{E}$, $\mathscr{P}(\mathscr{E})$ is the powerset of $\mathscr{E}$ and $|\mathscr{E}|$ is the cardinality of $\mathscr{E}$. We use $[\![1;k]\!]$ to denote the set of integers $\{1,\ldots,k\}$. We call C the set of all the candidates, R the set of rankings (permutations) on the candidates of C. We fix $m := |C|$. A ranking can be represented explicitly by the tuple that lists the candidates from the most to the least preferred; for instance, the tuple (b, c, a) denotes the ranking that ranks b in first position, c in second position and a in last position. Let $r \in R$, $a \in C$ and S be a non-empty subset of C. $r_S(a)$ denotes the rank of a in the restriction of ranking r that considers only the elements of S. In particular, we define $r(a) := r_C(a)$. $\text{top}_r(S)$ is the most preferred candidate by r among the candidates of S. We suppose that the preferences of every voter v over the candidates can be modelled by a ranking. Under the assumption of anonymity, we will consider voting situations [2] that we here model as multisets of rankings (since the same ranking may occur several times). Throughout the paper, we consider V, a multiset of $n \in \mathbb{N}$ rankings (representing voters).

The *availability distribution* is a probability distribution P on $\mathscr{P}(C)$ such that, for every $S \subseteq C$, $P(S)$ is the probability that the set of available candidates is exactly S. We use $\mathcal{P}_C$ to denote the set of probability distributions on $\mathscr{P}(C)$. The *disutility function* (DF), denoted by ρ, is an increasing mapping from $[\![1;m]\!]$ to $\mathbb{R}$ such that $\rho(i)$ measures how much a voter is unsatisfied by the item at the i-th position in his/her ranking. Given a ranking r, when the set S of available candidates is observed, the candidate $\text{top}_r(S)$ is declared the winner. The dissatisfaction felt by a voter v is then the difference between the disutility of $\text{top}_r(S)$ and that of $\text{top}_v(S)$. Finally, the total dissatisfaction is the sum of the dissatisfactions of all the voters in V. When we produce a ranking, the set of available

*This work was mainly conducted while at LIP6, Sorbonne Université

Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), U. Endriss, A. Nowé, F. Dignum, A. Lomuscio (eds.), May 3–7, 2021, Online.

candidates is not known; that is why we provide the ranking that minimises the total dissatisfaction in expectation over P.

3 EX ANTE VS. EX POST DISSATISFACTION

In the ex ante model, the dissatisfaction is computed using the positions of the candidates in the whole set of candidates C.

Definition 1. *Let $P \in \mathcal{P}_C$, ρ be a DF and $r \in R$. We define $\hat{\Delta}_{\rho,P}(V,r) \coloneqq \sum_{v \in V} \mathop{\mathbb{E}}_{S \sim P}[\rho(v(\text{top}_r(S))) - \rho(v(\text{top}_v(S)))]$ and the set of* optimal rankings *as $\hat{R}^*_{\rho,P}(V) \coloneqq \arg\min_{r \in R}[\hat{\Delta}_{\rho,P}(V,r)]$.*

Example 1. *Let $C = \{a,b,c\}$, $\rho = (0,1,2)$, $P \in \mathcal{P}_C$ be the uniform distribution: for all $S \subseteq C$, $P(S) = \frac{1}{8}$. Let V be composed of 4 voters voting according to the ranking $r' = (a,b,c)$ and 7 according to $r'' = (c,a,b)$. The following array displays, for all $r \in R$, from left to right, the value of $\text{top}_r(S)$ for every non-empty and non-singleton $S \subseteq C$, $\sum_{S \subseteq C} \rho(r'(\text{top}_r(S))$, $\sum_{S \subseteq C} \rho(r''(\text{top}_r(S))$ and, in the last column, $4\sum_{S\subseteq C}\rho(r'(\text{top}_r(S))+7\sum_{S\subseteq C}\rho(r''(\text{top}_r(S))=8[\hat{\Delta}_{\rho,P}(V,r)+\chi(V)]$ where $\chi(V) \coloneqq \mathop{\mathbb{E}}_{S \sim P}[4\rho(r'(\text{top}_{r'}(S))) + 7\rho(r''(\text{top}_{r''}(S)))]$ does not depend on r.*

r \ S	abc	ab	ac	bc	$\rho(r'(\text{top}_r(S))$	$\rho(r''(\text{top}_r(S))$	Σ
(a,b,c)	a	a	a	b	0 + 0 + 0 + 1	1 + 1 + 1 + 2	39
(a,c,b)	a	a	a	c	0 + 0 + 0 + 2	1 + 1 + 1 + 0	***29***
(b,a,c)	b	b	a	b	1 + 1 + 0 + 1	2 + 2 + 1 + 2	61
(b,c,a)	b	b	c	b	1 + 1 + 2 + 1	2 + 2 + 0 + 2	62
(c,a,b)	c	a	c	c	2 + 0 + 2 + 2	0 + 1 + 0 + 0	31
(c,b,a)	c	b	c	c	2 + 1 + 2 + 2	0 + 2 + 0 + 0	42

We deduce that the only optimal ranking is (a,c,b).

We provide a characterisation of the ex ante dissatisfaction rule which shows that, given a DF ρ, finding an optimal ranking boils down to using the scoring rule whose scoring function is $-\rho$ and that this result does not depend on the availability distribution P as long as P is non-null on pairs ($P(S) > 0$ if $|S| = 2$).

Theorem 1. *Let $P \in \mathcal{P}_C$ and ρ be a DF. The set of rankings where the candidates are in the increasing order of $a \in C \mapsto \sum_{v \in V} \rho(v(a))$ (there are several such rankings if some candidates have equal scores) is included in $\hat{R}^*_{\rho,P}(V)$. If P is non-null on pairs, the two sets are equal.*

In the ex post model, we assume that the disutility felt by a voter when a candidate is elected depends on its position *within* the set of the actually available candidates.

Definition 2. *Let $P \in \mathcal{P}_C$, ρ be a DF and $r \in R$. We define $\Delta_{\rho,P}(V,r) \coloneqq \sum_{v \in V} \mathop{\mathbb{E}}_{S \sim P}[\rho(v_S(\text{top}_r(S))) - \rho(1)]$ and the set of* optimal rankings *as $R^*_{\rho,P}(V) \coloneqq \arg\min_{r \in R}[\Delta_{\rho,P}(V,r)]$.*

Example 2. *Let us summarise the computations in the following array, in the same manner as in Example 1:*

r \ S	abc	ab	ac	bc	$\rho(r'_S(\text{top}_r(S))$	$\rho(r''_S(\text{top}_r(S))$	Σ
(a,b,c)	a	a	a	b	0 + 0 + 0 + 0	1 + 0 + 1 + 1	21
(a,c,b)	a	a	a	c	0 + 0 + 0 + 1	1 + 0 + 1 + 0	18
(b,a,c)	b	b	a	b	1 + 1 + 0 + 0	2 + 1 + 1 + 1	43
(b,c,a)	b	b	c	b	1 + 1 + 1 + 0	2 + 1 + 0 + 1	40
(c,a,b)	c	a	c	c	2 + 0 + 1 + 1	0 + 0 + 0 + 0	***16***
(c,b,a)	c	b	c	c	2 + 1 + 1 + 1	0 + 1 + 0 + 0	27

The ex post rule gives (c,a,b), and not (a,c,b) as the ex ante rule.

The ex post approach is also the approach considered in [1, 7] and, in particular, in [10] with the DF $(0,1,\ldots,1)$ which implicitly assumes that the best decision would be to follow plurality when the set of available candidates is revealed. A major advantage of plurality is the small quantity of information needed and, thus, a reduced cognitive load for voters. Since, in the unavailable candidate model, the complete rankings of preferences are needed anyway, generalising scoring rules and not only plurality via ex post rule does not require more information but provides a richer model.

4 ALGORITHMIC ANALYSIS

For the ex ante approach, it directly follows from our characterisation result that an optimal ranking in the ex ante dissatisfaction model can be found in *polynomial time* in m and n. For the ex post approach, the problem is more challenging.

We assume that the availability distribution is $P\colon S \subseteq C \mapsto p^{m-|S|}(1-p)^{|S|}$ for a $p \in]0;1[$. Note that, as far as $\rho(1) < \rho(2)$ (which we suppose true, otherwise ex post rule cannot distinguish some pairs of distinct rankings), ρ can be *normalised* so that $\rho(1) = 0$ and $\rho(2) = 1$ without changing the set of optimal rankings. Given $q \in [1;+\infty[$, we say that a normalised DF ρ is *q-sub-geometrical* if, for any $k \in [\![2;m]\!]$, $\rho(k) \leq q^{k-2}$.

Theorem 2. *Let ρ be a q-sub-geometrical DF, for a $q \in [1;+\infty[$. Let $\epsilon \in]0, \frac{1}{nm(m-1)+1}[$, $p \in [\max((1-\epsilon)^{\frac{1}{m-1}}, \frac{q-(1+\epsilon)^{\frac{1}{m-1}}}{q-1});1[$. Any ranking in $R^*_{\rho,P}(V)$ is also a Kemeny consensus.*

A direct consequence of Theorem 2 is that ex post rule is NP-hard. Taking inspiration from [10], we define a *MyopicTop* algorithm, but tailored to the ex post rule via our new notion of dominance:

Proposition 1. *If there is no ambiguity on ρ, we define $f\colon a \in C \mapsto \sum_{v \in V} \sum_{S \subseteq C, a \in S} P(S)\rho(v_S(a))$. Let $a \in C$. If, for all $b \in C \setminus \{a\}$, $(1+p)f(a) < (1-p)f(b)$, then, a is the first candidate of all optimal rankings. In this case, we call a the* dominant candidate.

The idea of the algorithm is as follows. We successively remove the dominant candidates from C and append them to the output ranking. When there is no more dominant candidate in the set of the remaining candidates, we choose and order K candidates so that they minimise their contribution to $\Delta_{\rho,P}(V,r)$ (this step is exponentially complex in K). Finally, the remaining candidates are ordered randomly.

Proposition 2. MyopicTop *algorithm runs in $O(nm^{\max(3,K+2)})$.*

It is possible to show that *MyopicTop* algorithm is a PTAS for the ex post dissatisfaction rule provided that ρ is normalised and bounded. Note that here, contrarily to Theorem 2, parameter p is assumed to be fixed and independent on n and m.

Theorem 3. *Let $\epsilon > 0$. Let $(\rho_m)_{m \in \mathbb{N}^*}$ be a family of normalised DF; we suppose that there exists a fixed $M \in \mathbb{R}^*_+$ such that, for all $m \in \mathbb{N}^*$, for all $i \in [\![1;m]\!]$, $\rho_m(i) \leq M$. Let $K \coloneqq \lceil \log_{\frac{1}{p}}(\frac{2M}{(1-p)^3\epsilon}) \rceil$. Let $m \in \mathbb{N}^*$. Consider an optimal $r^* \in R^*_{\rho_m,P}(V)$ and r the ranking obtained via the* MyopicTop *algorithm with inputs C, ρ, p, V, K. If $\Delta_{\rho_m,P}(V,r^*) = 0$, $\Delta_{\rho_m,P}(V,r) = 0$. Otherwise, $\frac{\Delta_{\rho_m,P}(V,r)}{\Delta_{\rho_m,P}(V,r^*)} \leq 1+\epsilon$.*

The assumption whereby the DF is bounded is quite reasonable in a context of preference aggregation in which the voter has more difficulty to discriminate between the alternatives as they go further in his/her preference ranking, explaining the convergence of the DF towards the upperbound which can be understood as the disutility of completely disliked alternatives.

REFERENCES

[1] Katherine A Baldiga and Jerry R Green. 2013. Assent-maximizing social choice. *Social Choice and Welfare* 40, 2 (2013), 439–460.

[2] Sven Berg and Dominique Lepelley. 1994. On probability models in voting theory. *Statistica Neerlandica* 48, 2 (1994), 133–146.

[3] Craig Boutilier, Jérôme Lang, Joel Oren, and Héctor Palacios. 2014. Robust winners and winner determination policies under candidate uncertainty. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*. AAAI, Québec.

[4] Yann Chevaleyre, Jérôme Lang, Nicolas Maudet, Jérôme Monnot, and Lirong Xia. 2012. New candidates welcome! Possible winners with respect to the addition of new candidates. *Mathematical Social Sciences* 64, 1 (2012), 74–88.

[5] Bhaskar Dutta, Matthew O Jackson, and Michel Le Breton. 2001. Strategic candidacy and voting procedures. *Econometrica* 69, 4 (2001), 1013–1037.

[6] Bhaskar Dutta, Matthew O Jackson, and Michel Le Breton. 2002. Voting by successive elimination and strategic candidacy. *Journal of Economic Theory* 103, 1 (2002), 190–218.

[7] Hugo Gilbert, Tom Portoleau, and Olivier Spanjaard. 2020. Beyond Pairwise Comparisons in Social Choice: A Setwise Kemeny Aggregation Problem. In *Thirty-Fourth AAAI Conference on Artificial Intelligence*. 1982–1989.

[8] Christian Klamler. 2008. A distance measure for choice functions. *Social Choice and Welfare* 30, 3 (2008), 419–425.

[9] Jérome Lang, Nicolas Maudet, and Maria Polukarov. 2013. New results on equilibria in strategic candidacy. In *International Symposium on Algorithmic Game Theory*. Springer, 13–25.

[10] Tyler Lu and Craig E Boutilier. 2010. The unavailable candidate model: a decision-theoretic view of social choice. In *Proceedings of the 11th ACM conference on Electronic commerce*. 263–274.

[11] Lihi Naamani-Dery, Meir Kalech, Lior Rokach, and Bracha Shapira. 2016. Reducing preference elicitation in group decision making. *Expert Systems with Applications* 61 (2016), 246–261.

[12] Joel Oren, Yuval Filmus, and Craig Boutilier. 2013. Efficient Vote Elicitation under Candidate Uncertainty.. In *IJCAI*. 309–316.