Two Framework Proposals and how they compare

COMSOC Project

Silvan Hungerbuehler

1 Proposal 1: Budgeted k-Borda

Basic Idea We construe the problem of giving a media recommendation to a group of people as the problem of finding that recommendation that maximizes readers' value while satisfying some budget side-constraint. Maximizing value (which could probably be recast as minimizing misrepresentation) will require coming up with some metric based on the profile and the recommendation; my preliminary suggestion for this is to use the Borda score.

Max: Another way to frame the problem: We are given a profile of preference orders. The interpretation is that the preferences are "revealed" by an individual recommendation algorithm using readers' past behaviour, likes and other data. Based on this profile we want to give a group recommendation assuming that each reader will read the whole recommendation. The latter assumption receives the strongest justification if a) it adheres to a budget (of e.g. time, effort) set by the readers and b) minimizes regret/complaint the readers have when comparing the selection to their individual recommendation.

Framework We have a set of *news items* $A = \{a_1, ..., a_m\}$, each having a specific *cost* $C : A \to \mathbb{R}$, a set of *recommended items* $W \subseteq A$, a set of *consumers* $N = \{n_1, ..., n_n\}$, a *profile of preferences* over the set of items $\mathbf{R} \in \mathscr{L}^n$ and a *budget* $B \in \mathbb{R}_{\geq 0}$....
Further, there is a *value function*, akin to a Borda vector, indicating how much a consumer values an option in her ballot amongst the recommended items. It takes as an input the consumer, the profile and an element of A, $V : \mathscr{L}^n \times N \times A \to \mathbb{R}$. It is akin to a Borda in the concrete case where V outputs the value $m-1$ for all consumers' candidates in the top position, $m-2$ for the candidates in the second position and so forth. This is very general and is perhaps more conventiently expressed with vector notation, but I could not figure out how to express the maximization problem that way, so I hope the idea is clear.

Haukur: I find the value definition a bit too general. It allows for a voter to base her utilities on other people's perferences and I see this as opening of a plethora of possibilities and expressibility which we do not need. We could rather assume that consumers' utilities are independent of each other and only based on the item, not considering what other itemes are in W *intrinsic utility*. Essentially, each consumer i assigns a value to item a_j. For the specific Borda case, $u_i(a_j) = m-1$ if a_j is the item i likes the most. Furthermore, we take the *utilitarian* view and measure the social welfare as the sum of each agent's percieved utility. We might want to consider *rank* *based utilities* in which consumers evaluate W with respect to the order each item in W appears in their preference order, see formulation later.

Discussion with Max & Greg: We should not consider *rank based utilities* because we should assume that all readers read everything. Therefore we should use K-borda for this problem.

Max: If I understand correctly, in both Silvan's and Haukur's picture it is possible for voters to have different scoring vectors, e.g. a might have a Borda vector over their profile and b a k-approval vector. While this gives a lot of expressibility, the problem I have with it is that it essentially requires utilities as *input*. A preference order plus a scoring vector is nothing else than a utility. I'm fine with constructing pseudo-utilities for technical reasons in a generic way by e.g. always using the same Borda-vector. However, using individual scoring vectors would require information about the individual voters' utilities which would go beyond the realm of voting rules. If we were to go this way, I think there is no reason why we don't directly consider a profiule of utilities instead of a profile of preferences plus individual scoring vectors - no need for the value function. However, a) this requires much more information in the input and b) would not really be about voting rules anymore. Therefore I would prefer to stick with a generic scoring vector. An alternative would be to use an impartial culture assumption of sorts, e.g. randomly generate scoring vectors with monotonely descending values.

Regarding the voting rule. If we use a fixed k then this is indeed k-Borda if I'm not mistaken. However, in both Silvan's and Haukur's framework the constraint is not some fixed k but a fixed budget B. So the voting procedure is rather "maximisize the winning set's (Borda) score under a budget constraint".

What we try to maximize is the sum of all the consumers' values by choosing W (of course this can be recast as a minimization problem by adjusting V). For each consumer we only count the value from the news items that are actually in the recommended set:

$$\max_{W} \sum_{j=1}^{n} \sum_{i=1}^{m} \mathbb{1}[a_i \in W] V(\mathbf{R}, n_j, a_i) \tag{1}$$

Of course, Equation 1 is trivially solved by $W = A$. But the interest in solving it comes from adding the budget constraint.

$$\max_{W} \sum_{j=1}^{n} \sum_{i=1}^{m} \mathbb{1}[a_i \in W] V(\mathbf{R}, n_j, a_i) \text{ subject to } \sum_{a_i \in W} C(a_i) \leq B \tag{2}$$

Haukur: Different problem statement. I use $\boldsymbol{w} = [w_1, w_2, ..., w_m]$ to represent what items of A are in the winning set.

$$\max \sum_{j=1}^{n} \sum_{i=1}^{m} u_i(a_j) \cdot w_j \tag{3}$$

Subject to

$$w_j \in \{0, 1\}, 1 \leq j \leq |A|, j \in \mathbf{N} \tag{4}$$

$$\sum_{i=1}^{m} c(a_j) \cdot w_j \leq B \tag{5}$$

Do we want to assume that all consumers will read all the articles? [Yes!] If not, we might want to consider that each consumer will mostlikely read her favorite article and is less likely to read the second favorite article and therefore will gain less utility by doing so. We need to add more constraints and a different problem statement to capture this. Now I use $w_{i,j,k} = 1$ iff item a_j is kth prefered item for i and is in W.

$$\max \sum_{j=1}^{n} \sum_{i=1}^{m} \sum_{k=1}^{|W|} u_i(a_j) \cdot w_{i,j,k} \tag{6}$$

Subject to

$$w_j \in \{0,1\}, 1 \leq j \leq |A|, j \in \mathbf{N} \tag{7}$$

$$w_{i,j,k} \in \{0,1\}, 1 \leq i \leq |N|, 1 \leq j \leq |A|, 1 \leq k \leq |W|, j,i,k \in \mathbf{N} \tag{8}$$

$$\sum_{i=1}^{m} c(a_j) \cdot w_j \leq B \tag{9}$$

$$w_{i,j,k} \leq w_i \tag{10}$$

Captures if $w_{i,j,k} = 1$ then $w_i = 1$

$$\sum_{j=1}^{m} w_{i,j,k} = 1 \tag{11}$$

For each consumer and rank pair there is only one item.

$$\sum_{k=1}^{|W|} w_{i,j,k} \leq 1 \tag{12}$$

For each consumer and item pair it is only ranked at most once.

$$\sum_{j=1}^{m} u_i(a_j) \cdot w_{i,j,k} \geq \sum_{j=1}^{m} u_i(a_j) \cdot w_{i,j,k+1} \tag{13}$$

We only consider cases in which $w_{i,j,k}$ is in acsending order according to the consumer's utilities. If we use this formalization we capture the

2 Proposal 2: θ-Smith rule

Investigations Max: I investigated Condorcet winning sets and it turns out they are probably not useful for us. Turns out a Condorcet winning set is defined *disjunctively*, i.e. a set W is Condorcet-winning if for every alternative $a \in X \setminus W$, there is majority of voters such that every voter in the majority prefers *some* element of W over a. This requirement is very weak s.t. most elections have already multiple Condorcet Winning sets of size 2 and it is not even known whether any election exist with a smallest Condorcet winning set of size greater than 3.

Thus the criterion is usually strenghtened by not just requiring majority but a higher threshold θ. With a high enough threshold the rule "smallest θ-winning set" becomes more decisive.

I have a hard time of getting a good intuition what kind of properties such a rule has and whether they are desirable in our context. Maybe some of you have a better intuition?

Possibly of greater interest is the concept of Condorcet Committee. Condorcet Committees are defined *conjunctively*, i.e. W is a Condorcet committee if every $w \in W$

beats every $a \in X \setminus W$ in pairwise majority contests. The smallest Condorcet Committee given a profile is called a Smith set. This already implies the problem that like a Condorcet Winner, a Condorcet Committee of a given size k needn't always exist. Thus the voting rule is usually extended to some k-Condorcet-Committee consistent rule.

Basic Idea This idea currently does not work! One could use Condorcet Committees but go a different avenue. Analogously to before,we relax the majority requirement calling the committees obtained θ-commitee (e.g. every w in the 0.3-commitee W gets at least 30% in pairwise majority contests against every $a \in X \setminus W$). Likewise, we generalise the notion of a Smith set to θ-Smith sets. Instead of requiring a fixed k, we again set a budget constraint. We then choose the *smallest* θ for which the θ-Smith set is still compatible with the budget.

Framework As before, we have a set of *news items* $A = \{a_1, ..., a_m\}$, each having a specific *cost* $C : A \to \mathbb{R}$, a set of *recommended items* $W \subseteq A$, a set of *consumers* $N = \{n_1, ..., n_n\}$, a *profile of preferences* over the set of items $\mathbf{R} \in \mathscr{L}^n$ and a *budget* $B \in \mathbb{R}_{\geq 0}$.

W is a θ-*committee* if for every $w \in W$, $a \in X \setminus W$ $|\{n_i : w \succ_i z\}| > \theta N$. For a given θ, let $cond_{\theta,\mathbf{R}} \subseteq 2^A$ be the set of all θ-committees of a given profile $\mathbf{R}$. Given a set of sets B denote by min(B) the function that outputs the element of B with the smallest cardinality (together with some tie breaking rule). Then our voting rule is characterized by $W = min(cond_{\theta,\mathbf{R}})$ and the following optimization problem:

$$\min \theta \text{ subject to } \sum_{a_i \in W} c(a_i) \leq B$$

3 Proposal III: Least-θ-rule

Basic Idea Instead of a Condorcet Committee, we want to elect the *set of θ-winners* Θ, that is the set of all alternatives that win at least θ% in all majority contests. Given the budget constraint, we choose the smallest θ that is still compatible with the budget.

Framework We define Θ as follows: let $a : b$ denote the majority contest between a and b and $\#_a(a : b) \in [0, 1]$ the share of the vote that a wins against b in that contest. Then $\Theta = \{a \in A : \text{ for all } b \in A, \#_a(a : b) \geq \theta\}$. Then W=$\Theta$ subject to the following optimization problem:

$$\min \theta \text{ subject to } \sum_{a_i \in W} c(a_i) \leq B$$

4 Comparison

Discussion with Haukur and Greg yielded the idea of a comparison of a θ-Smith-set-consistent approach to a PSR-based approach with respect to certain desirable properties of the winner sets (or "axioms"). Here are three (which are insufficiently formalized and of course debatable):

- **Minimal Regret/Maximal utility**: the voting rule minimisizes regret/maximizes utility.
- **Equality**: readers' utility is within a certain intervall.
- θN**-minority consistency**: If a minority of θN-readers ranks an item first, then the item will be in W.

In additions the axioms from Elkind et. al. 2014 would be of interest.

Questions: Are these properties indeed desirable? How bad is the loss in utility for the θ-Smith rule compared to the budgeted k-Borda rule? How much worse is budgeted k-Borda compared to the θ-Smith rule wrt θN-minority consistency? How do the two rules fare in terms of equality?

Possible starting point: Diss and Doghmi 2016 have studied how likely it is for k-Borda to pick the k-Condorcet Committee. I have put the paper in the Bibtex file.