Budgeted Social Choice: A Framework for Multiple Recommendations in Consensus Decision Making

Tyler Lu and Craig Boutilier

Abstract

We develop a new framework for social choice problems, *budgeted social choice*, in which a *limited* number of alternatives can be recommended/prescribed to a population of agents. This limit is determined by some form of budget. Such problems naturally arise in a variety of contexts. Our model is general, spanning the continuum from pure consensus decisions (i.e., standard social choice) to fully personalized recommendation. Our results show that standard rank aggregation rules are not appropriate for such tasks and that good solutions typically involve picking diverse alternatives tailored to different agent types. The corresponding optimization problems are shown to be NP-complete, but we develop fast greedy algorithms with some theoretical guarantees. Experimental results on real-world datasets (APA election and sushi) show some interesting patterns and the prove the effectiveness of our greedy algorithms.

1 Introduction

Social choice has received considerable attention in AI and computer science in recent years [10, 13, 7]. This is in part due to technological advances that have facilitated an explosion in the availability of (sometimes implicit) ranking or preference data. Users can, with increasing ease, rate, compare or rank products (e.g., movies, consumer goods, neighborhoods) and information (e.g., clicking on search responses or ads, linking to data sources in social media). This has allowed a great degree of personalization in product recommendation and information provision.

Despite this trend, tailoring the alternatives presented or recommended to specific users can be difficult for any of a number of reasons, among them privacy concerns (actual or perceived), scarce data, or the infeasibility of complete personalization. For example, decisions regarding certain types of public projects (such as highway placement, or park design) may force the choice of a single option: one cannot build different projects to meet the desires of different individuals. Similarly, a company designing a product to meet consumer demand must find a single product that maximizes consumer satisfaction across its target market (assuming sufficient correlation between satisfaction and revenue/profit). In such settings, a single "consensus" recommendation must be made for the population as a whole. If such consensus recommendations are made in a way that is sensitive to the preferences of individuals, we land squarely in the realm of social choice.

There is, of course, a middle ground between pure personalization and pure consensus recommendation. For example, suppose the company can configure its manufacturing facility to produce three variants of the product in question. Then its aim should be to determine three products that *jointly* maximize consumer satisfaction. In the case of public projects, perhaps a small number of projects can be chosen. In domains like web search, if one has insufficient data about an individual making a query (or is reluctant to use it because of privacy concerns), a small number of responses can be presented if browser "real estate" is limited. In the design of pension plan options, there are many reasons to limit the number of offerings available to encourage meaningful choice. In these and numerous other examples, we fall somewhere between making a single consensus recommendation and making fully personalized recommendations for individuals. Some (perhaps implicit) aggregation of users must take place—we cannot offer fully personalized offerings to each individual—placing us in the realm of social choice; but at the same time, we have an opportunity to do some tailoring of the decisions to the preferences of the aggregated groups, and indeed, make choices about the precise form of this aggregation to optimize some social choice function.

In this paper, we develop a general model for just such settings. We call the problem at hand one of *budgeted social choice*. Unlike the usual social choice models, in which a single outcome is selected (or single consensus ranking determined), we allow for the possibility that more than one option can be offered, and assume that each user will *benefit from the best option, according to her own preferences, among those presented*. However, the number of options offered is constrained by a *budget*; this is the key factor that prevents us from exploiting pure personalization to meet the desires of individual users. This budget can take a variety of forms, and we explore several of them in this work. The budget could be a strict limit on the *number* of options (e.g., at most three products can be manufactured, or at most 10 web links can be presented on a page), or on *their cost* (e.g., the total expenditure on city parks cannot exceed $3M). We can also adopt a more nuanced perspective in which the cost of allowing additional options is traded off against the benefit to the target population (e.g., add a fourth product option if increase in consumer satisfaction outweighs the cost of a fourth production line; or extend the city parks budget if increase in social welfare is sufficiently high). Finally, we can consider settings in which the budget is not just a function of the options "created," but also of their overall usage or uptake in the population. Our general framework allows for a fixed charge (e.g., configuring and staffing an assembly line) and per-unit cost (e.g., the marginal cost of producing a unit of product for a specific individual).

Though the motivations are different, multiple-winner models in voting theory [4, 20] can be viewed as an instance of our model. In such systems, the goal is to determine a collection of candidates (e.g., a parliament) that best represents the "collective interests" of the voters (e.g., based on principles of proportional representation). Indeed, our "limited choice" model with Borda scoring corresponds directly to Chamberlin and Courant's [4] proportional representation scheme; in this way, our budgeted choice model can be used to motivate the application of such proportional models to ranking and recommendation, under certain assumptions. Also related is the combinatorial public project problem [19] where given each agent's valuation over all subsets of alternatives, a limited number of alternatives must be chosen for everyone. The focus is more on the tension between approximating social welfare and incentivizing truthfulness (requiring payments from agents).

We begin by outlining a simple model of budgeted social choice in which there is a strict limit K on the number of candidates that can be made available. We do this to illustrate the general principles and intuitions underlying our approach and draw connection to proportional representation schemes. We show that for various social choice objectives, computing the optimal set of K candidates for a set of preferences in this *limited choice model* is NP-hard. However, the induced objective is submodular, and a simple greedy algorithm produces candidate sets whose deviation from optimal is bounded. Computational experiments on various preference data sets show that the greedy algorithm is, in fact, very close to optimal in practice.

We then present our general model in which adding alternatives to the available set is costly (allowing both fixed and per-unit charges) and subject to some form of budget. The limited choice model is a special case of this *costly choice model*. The costly choice model with only fixed charges remains submodular, but when per-unit costs are included, submodularity vanishes. We develop an integer programming formulation of the general optimization problem (which applies directly to the limited choice model). We again provide a greedy heuristic algorithm for solving the general problem which runs in polynomial time. Computational experiments verify its efficacy in practice, but we have no theoretical bounds on its performance currently.

2 Background

We first review some basic concepts from social choice before defining the class of budgeted social choice problems (see [11] for further background). We assume a set of *agents* (or *voters*) $N = \{1, \ldots, n\}$ and a set of *alternatives* (or *candidates*) $A = \{a_1, \ldots, a_m\}$. Let Γ_A be the set of *rankings* (or *votes*) over A (i.e., permutations over A). Alternatives can represent any outcome space over which the voters have preferences (e.g., product configurations, restaurant dishes, candidates for office, public projects, etc.) and for which a single collective choice must be made. Agent ℓ's preferences are represented by a ranking $v_\ell \in \Gamma_A$, where ℓ prefers a_i to a_j, denoted as $a_i \succ_{v_\ell} a_j$, if

$v_\ell(a_i) < v_\ell(a_j)$. We refer to a collection of votes $V = (v_1, \ldots, v_n) \in \Gamma_A^n$ as a *preference profile*.

Given a preference profile, there are two main problems in social choice. The first is selecting a *consensus alternative*, requiring the design of a *social choice function* $f : \Gamma_A^n \to A$ which selects a "winner" given voter rankings/votes. The second is selecting a *consensus ranking* [2], requiring a *rank aggregation function* $f : \Gamma_A^n \to \Gamma_A$. The consensus ranking can be used for many purposes; e.g., the top-ranked alternative can be taken as the consensus winner, or we might select the top k alternatives in the consensus ranking in settings where multiple candidates can be chosen (say, parliamentary seats, or web search results [10]). *Plurality* is the simplest, most common approach for consensus alternatives: the alternative with the greatest number of "first place votes" wins (various tie-breaking schemes can be adopted). However, plurality fails to account for a voter's relative preferences for any alternative other than its top ranked (assuming sincere voting). Other schemes, e.g., *Borda count* or *single transferable vote*, produce winners that are more sensitive to relative preferences. Among schemes that produce consensus rankings, the *Borda ranking* [8] and the *Kemeny consensus* [15] are especially popular.

Definition 1. *Given a ranking v, the* Borda count *of alternative a is $\beta(a, v) = m - v(a)$. The* Borda count *of a relative to preference profile V is $\beta(a, V) = \sum_{v \in V} \beta(a, v)$. A* Borda ranking *$r^*_\beta = r^*_\beta(V)$ is any ranking that orders alternatives from highest to lowest Borda count.*

One can generalize the Borda count by assigning arbitrary scores to the rank positions:

Definition 2. *A* positional scoring function (PSF) *$\alpha : \{1, \ldots, m\} \to \mathbb{R}_{\geq 0}$ maps ranks onto scores s.t. $\alpha(1) \geq \cdots \geq \alpha(m) \geq 0$. Given a ranking v_ℓ and alternative a, let $\alpha_\ell(a) = \alpha(v_\ell(a))$. The* α-score *of a relative to profile V is $\alpha(a, V) = \sum_{v_\ell \in V} \alpha_\ell(a)$. An* α-ranking *$r^*_\alpha = r^*_\alpha(V)$ is any ranking that orders alternatives from highest to lowest α-score.*

Definition 3. *Let $\mathbf{1}$ be the indicator function,* sgn *the sign function and r, v two rankings. The* Kendall-tau *metric is $\tau(r, v) = \sum_{1 \leq i < j \leq m} \mathbf{1}[\operatorname{sgn}[(v(a_i) - v(a_j))(r(a_i) - r(a_j))] < 0]$. Given a profile V, the* Kemeny cost *of a ranking r is $\kappa(r, V) = \sum_{v_\ell \in V} \tau(r, v_\ell)$. The* Kemeny consensus *is any ranking $r^*_\kappa = r^*_\kappa(V)$ that minimizes the Kemeny cost.*

Intuitively, Kendall-tau distance measures the number of pairwise relative misorderings between an output ranking r and a vote v, while the Kemeny consensus minimizes the total number of such misorderings across profile V. While positional scoring is easy to implement, much work in computational social choice has focused on NP-hard schemes like Kemeny [10, 3].

Rank aggregation has interesting connections to work on *rank learning*, much of which concerns aggregating (possibly noisy) preference information from agents into full preference rankings. For example, Cohen et al. [6] focus on learning rankings from (multiple user) pairwise comparison data, while *label ranking* [13] considers constructing personalized rankings from votes. Often unanalyzed is why specific rank aggregations should be chosen for particular settings such as these. One can think of some schemes as a maximum likelihood estimator of some underlying objective ranking (e.g., for Kemeny [22] and positional scoring rules [7]).

3 The Limited Choice Model

While the use of social choice techniques in applications like web search and recommender systems is increasingly common, the motivations for producing consensus recommendations for users with different preferences often varies. Consider, for instance, the motivation for "budgeted" consensus recommendation discussed in our introduction. If a decision maker can provide a limited set of K choices to a population of users to best satisfy their preferences, methods like Kemeny, Borda, etc. *could* be used to produce an aggregate ranking from which the top K alternatives are taken. However, there is little rationale for doing so without a deeper analysis of what it means to "satisfy" the preferences of the user population. In the spirit of our recent work on rank aggregation [17], we develop a precise decision-theoretic formulation of the budgeted social choice problem. Rather than applying existing social choice schemes directly, we derive optimal consensus decisions from decision-theoretic principles and show how these differ (and relate to) classic aggregation rules.

We first introduce the *limited choice problem*, a simple version of budgeted social choice in which one must choose a slate of K alternatives that maximizes some notion of total satisfaction among a group of agents. We develop the more general budgeted model in the next section. Assume a set of n voters with preferences over alternatives A as above. Rather than selecting a single consensus alternative, a decision maker is allowed to recommend K alternatives. Each voter realizes benefit commensurate with its most preferred alternative among the K recommended. For example, a company may be limited to offering K products to its target market, where the products are substitutes (so no consumer will use more than one); or a municipality may have budget for K new parks and citizens draw enjoyment from their most preferred park.

While our goal is to find the best set of K alternatives, the formalization of this model depends on two key choices: how voter satisfaction with a slate is measured; and how we measure social welfare. Our general framework can accommodate many measures of utility and social welfare, but for concreteness we focus on (a) positional scoring (such as Borda) to quantify voter satisfaction; and (b) the sum of such voter "utilities" as our social welfare metric. In other words, our aim is to find a slate of size K that maximizes the sum of the positional scores of each voter's most preferred candidate in the slate:

Definition 4. *Given alternatives A, preference profile V, and PSF α, a K-*recommendation set *is any set of alternatives $\Phi \subseteq A$ of size K. The α-*score *of Φ is:*

$$S_\alpha(\Phi, V) = \sum_{\ell \in N} \max_{a \in \Phi} \alpha_\ell(a) \ . \tag{1}$$

The optimal K-recommendation set *w.r.t. α is:*

$$\Phi^*_\alpha = \operatorname*{argmax}_{|\Phi|=K} S_\alpha(\Phi, V) \ . \tag{2}$$

We use $S_\alpha(\Phi, v)$ to denote the score w.r.t. a single vote/ranking v. We drop the subscript α from S_α when it is evident from the context, and use S_β to denote the special case of Borda scoring.

The objective in Eq. 2 is identical to the Chamberlin and Courant [4] scheme of proportional representation and results for that scheme apply directly to this variant of the limited choice model, as we discuss below. While we focus on total positional scoring as our optimization criterion, the general budgeted framework allows other measures of utility and social desiderata. For example, we can use maximin-fairness (w.r.t. positional scoring) encoded as:

$$\Phi^*_{\text{fair}} = \operatorname*{argmax}_{|\Phi|=K} \min_{\ell \in N} S_\alpha(\Phi, v_\ell) \ . \tag{3}$$

Setting $\alpha(i) = \mathbf{1}[i = 1]$ corresponds to a binary satisfaction measure in which a voter is satisfied with Φ only if its top alternative is made available. In this case, the optimal Φ^*_α corresponds to selecting the K alternatives with the highest "plurality" score (i.e., greatest number of first-place "votes"). However, choosing the top K candidates from a consensus ranking using positional scoring is, in general, not appropriate. For any ranking r, let $r|K$ denote the K top-ranked alternatives in r. The Borda ranking r^*_β can produce slates $r^*_\beta|K$ that are a factor of 2 from optimal using our limited-choice measure, while the α-ranking for arbitrary PSFs can be as much as a factor of K from optimal.

Proposition 5. *For any K we have: (a) $\inf_{(m,n,V)} \frac{S_\beta(r^*_\beta|K,V)}{S_\beta(\Phi^*,V)} = 1/2$; and (b) $\inf_{(\alpha,m,n,V)} \frac{S_\alpha(r^*_\alpha|K,V)}{S_\alpha(\Phi^*,V)} \leq 1/K$.*

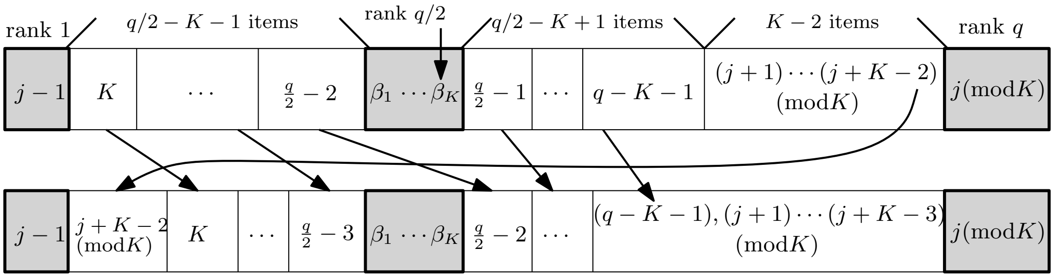

Fig. 1: Example showing that $r^*_\beta|K$ can be factor of 2 worse than optimal. Assume q items $\{0,1,\ldots,q-K-1,\beta_1,\ldots,\beta_K\}$, and $n=K(q-K-2)$ votes. The votes are divided into K blocks, each containing $q-K-2$ votes. For each block $j\le K$, item $j-1$ is always the top alternative in each vote, and item $j \pmod K$ is the worst. This means the optimal recommendation set is $\Phi^*=\{0,\ldots,K-1\}$, with $S_\beta(\Phi^*,V)=(q-1)n$. The jth block of votes has a structure illustrated in the figure, with two example votes shown: the items j and $j \pmod K$ are fixed in the top/bottom spots and items $\beta_1,\ldots,\beta_K$ are also fixed in positions $q/2-K+1,\ldots,q/2$. (Fixed items are shaded.) The remaining items are arranged in the other positions in the first vote (the non-shaded positions). Starting with one such arrangement (e.g., the top vote in the figure), each candidate is "rotated downward" one non-shaded position (with wrap around) to produce the next vote in the block. This is repeated until $q-K-2$ votes are constructed for block j (i.e., one vote for each non-shaded position). Thus, any non-fixed item occupies each non-shaded rank position in exactly one vote in this block j. Thus, the *average* score of a non-shaded item is $\sum_{i\in[q-2]\setminus\{q/2,\ldots,q/2+K-1\}} i = \frac{-q^2+3q-2+qK+K^2-K}{-2q+2K+4} < q/2$ (whenever $q>K+2$, which always holds). Hence the average score of any item in $\{K,\ldots,q-K-1\}$ (which occupy only unshaded positions in all blocks) across all blocks is less than $q/2$. Also observe that the average score of any item in Φ^* is less than $q/2$: item $j-1$ has score $q-1$ in block j but has score 0 in block $j-2 \pmod K$ (giving average $(q-1)/2$ in these two blocks) and has average less than $q/2$ across all other blocks (since it is an unshaded item in those blocks). But the average score of β_i is at least $q/2$ (since its position is fixed in all blocks). Hence the top K items of the Borda ranking r^*_β are $\beta_1,\ldots,\beta_K$. But $S_\beta(r^*_\beta|K,V)=(q/2+K-1)n$, so $S(r^*_\beta|K,V)/S(\Phi^*,V)=(q/2+K-1)/(q-1)$, which approaches $1/2$ from above as $q\to\infty$.

Proof Sketch. (a) To obtain a lower bound, we note that the total Borda score of all alternatives is $\sum_{a\in A}\beta(a,V)=n(0+1+2+\cdots+m-1)=nm(m-1)/2$. The item a^*_β with the highest Borda count must have a count at least the average, over the alternatives, $nm(m-1)/2/m=n(m-1)/2$. Since a^*_β is the highest-ranked element in r^*_β, we have $S_\beta(r^*_\beta|K,V)\ge n(m-1)/2$. By contrast, the score of the optimal set Φ^* is at most $n(m-1)$. Hence $r^*_\beta|K$ has score that is no worse than a factor of $[n(m-1)/2]/[n(m-1)]=1/2$ from optimal. We demonstrate an upper bound realizing this worst-case error using the example described in Fig. 1.

(b) An upper bound can be demonstrated using an example somewhat similar in spirit to that for the Borda count as in (a); we omit it due to lack of space. It remains open whether $r^*_\alpha|K$ can indeed be worse than a factor of K from optimal. $\square$

These results illustrate that care must be taken in the application of rank aggregation methods to novel social choice problems. In our limited choice setting, the use of positional scoring rules (e.g., Borda) to determine the K most "popular" alternatives can perform extremely poorly. Intuitively, the optimal slate appeals to the *diversity* of the agent preferences in a way that is not captured by "top K" methods. Indeed, this is one of the motivations for the proportional schemes [4, 20]. More importantly, the underlying preference aggregation scheme is defined relative to an explicitly articulated decision criterion. We defer a detailed discussion for lack of space, but we note that STV, often used for proportional representation [21] can perform poorly w.r.t. our criterion as well. Specifically, we can show that the slate produced by STV can be a factor of 2 worse than optimal.

The examples above suggest that determining optimal recommendation sets in the limited choice model may be computationally difficult. This is the case: the problem is NP-complete even for in the specific case of determining voter satisfaction using Borda scoring:[1]

[1]The NP-hardness of a *variant* of the Chamberlin and Courant [4] proportional scheme is shown in [21], but the variant allows for arbitrary misrepresentation scores. The added flexibility in the reduction used means that it does *not* imply the

Theorem 6. *Given preference profile V, integer $K \geq 1$, and $t \geq 0$, deciding whether there exists a K-recommendation set Φ with (Borda) score $S_\beta(\Phi, V) \geq t$ is NP-complete.*

Proof Sketch. Membership in NP is easily verified. For hardness, we reduce an arbitrary *hitting set* instance to our problem: given $E = \{e_1, \ldots, e_p\}$, a set $\{B_1, \ldots, B_q\}$ of subsets of E, and integer $h \geq 1$, is there a $C \subseteq E$ of size at most K such that $\forall i \in \{1, \ldots, q\}, C \cap B \neq \emptyset$? We reduce this to our decision problem, with voters $N = \{1, \ldots, q\}$, alternatives $A = E \cup \{z_{ij} : i \in [q], j \in [\sum_{\ell=1}^{q} |B_\ell|]\}$, $m = |A|$, and $t = qm - \sum_{\ell=1}^{q} |B_\ell|$. Each voter ℓ has a preference ordering with elements in B_ℓ at the top (in arbitrary order), followed by $z_{\ell 1} z_{\ell 2} \cdots z_{\ell t}$, and with remaining alternatives $A \backslash B_\ell$ (in arbitrary order) at the bottom.

Any positive hitting set instance (say, with certificate C) corresponds to positive instance for in our problem. We simply take $\Phi = C$, and have $S_\beta(\Phi, V) \geq \sum_{\ell=1}^{q} m - |B_\ell|$ since, for each voter ℓ, there is an $e \in C$ that is in B_ℓ by definition of a hitting set. Summing the scores of the most preferred alternatives, $\max_{a \in \Phi} m - v_\ell(a) \geq m - |B_\ell|$, over all voters, gives $S_\beta(\Phi, V) \geq t$.

Suppose we have a negative hitting set instance. Consider any Φ that maximizes $S_\beta(\cdot, V)$. If Φ does not hit some B_ℓ then let $a' = \operatorname{argmin}_{a \in \Phi} v_\ell(a)$. If $a' \neq z_{\ell j}$ for any j then $m - v_\ell(a') < m - \sum_{\ell=1}^{q} |B_\ell|$ and $S_\beta(\Phi, V) < t$. Otherwise $a' = z_{\ell 1}$; but this implies that we can replace each such $z_{\ell 1} \in \Phi$ by some $b \in B_\ell$, which further implies that Φ hits *every* such B_ℓ and is thus a hitting set solution (contradiction). Hence, $S_\beta(\Phi, V) < t$. ☐

We can formulate this NP-hard problem as an integer program (IP) with $m(n+1)$ variables and $1 + mn + n$ constraints. We note that [20] provide a similar IP for the Chamberlin and Courant proportional scheme. Let $x_i \in \{0, 1\}$, $i \leq m$ denote whether alternative a_i appears in the recommendation set Φ, and let $y_{\ell i} \in \{0, 1\}$, $\ell \leq n, i \leq m$ denote whether a_i is the most preferred element in Φ for voter ℓ. We then have:

$$\max_{x_i, y_{\ell i}} \quad \sum_{\ell \in N} \sum_{i=1}^{m} \alpha_\ell(a_i) \cdot y_{\ell i} \tag{4}$$

$$\text{subject to} \quad \sum_{i=1}^{m} x_i \leq K, \tag{5}$$

$$y_{\ell i} \leq x_i, \qquad \forall \ell \leq n, i \leq m \tag{6}$$

$$\sum_{i=1}^{m} y_{\ell i} = 1, \qquad \forall \ell \leq n. \tag{7}$$

Constraint (5) limits the slate to at most K alternatives (a optimal set of size less than K can be expanded arbitrarily to size K, since score is nondecreasing in size). Constraints (6) and (7) ensure voters benefit only from alternatives in Φ, and benefit from exactly *one* such element. The objective is simply $S_\alpha(\Phi, V)$. An optimal solution will always have $y_{\ell i} = 1$ where a_i is ℓ's most preferred alternative in the set defined by the x_i.

The IP may not scale to large problems. Fortunately, this is a constrained submodular maximization, which admits a simple greedy algorithm with approximation guarantees [18].

Algorithm Greedy**.** We receive inputs α, V and integer $K > 0$. Initially $\Phi_0 \leftarrow \emptyset$. We then update Φ iteratively K times, each time updating the recommendation set by adding the item that increases score the most, i.e., $\Phi_i \leftarrow \Phi_{i-1} \cup \{\operatorname{argmax}_{a \in A} S(\Phi_{i-1} \cup \{a\}, V)\}$. We output Φ_K.

Theorem 7. *For any given preference profile V, the function $S(\cdot, V)$ defined over 2^A, with $S(\emptyset, V) = 0$, is submodular and non-decreasing. Consequently, the constrained maximization of Eq. (2) can be approximated within a factor of $1 - \frac{1}{e}$ by* Greedy*. That is, $\frac{S(\mathsf{Greedy}, V)}{S(\Phi^*, V)} \geq 1 - \frac{1}{e}$.*

NP-hardness of our limited choice model.

Proof. Let $\Phi \subseteq \Phi' \subseteq A$, $a \in A$ and $v \in V$. It is clear that $S(\Phi, v) \leq S(\Phi', v)$. Since $S(\cdot, v)$ is non-decreasing for any vote v v, it is non-decreasing over profiles V, i.e., $S(\Phi', V) \geq S(\Phi, V)$.

If a is v's strictly most preferred alternative among those in Φ', then $S(\Phi \cup \{a\}, v) = S(\Phi' \cup \{a\}, v) = \alpha(v(a))$. Since $S(\Phi, v) \leq S(\Phi', v)$, this implies $S(\Phi \cup \{a\}, v) - S(\Phi, v) \geq S(\Phi' \cup \{a\}, v) - S(\Phi', v)$. If a is not strictly most preferred by v within the set Φ', then $S(\Phi' \cup \{a\}, v) = S(\Phi', v)$, hence $S(\Phi' \cup \{a\}, v) - S(\Phi', v) = 0$. Since $S(\Phi \cup \{a\}, v) \geq S(\Phi, v)$, again we have $S(\Phi \cup \{a\}, v) - S(\Phi, v) \geq S(\Phi' \cup \{a\}, v) - S(\Phi', v)$. This implies, by definition, the submodularity of $S(\cdot, v)$ for any vote v. Since the sum of submodular functions is also submodular, $S(\cdot, V)$ is submodular for profiles V. The $1 - \frac{1}{e}$ approximation ratio follows from [18]. □

Constructing a slate of K alternatives maximizing total positional score is similar to the K-*medians problem*, where at most K facilities (alternatives) need to be located to serve their nearest customers (voters) while minimizing the total distance between customers and their nearest facility. Distance corresponds to voter dissatisfaction with alternatives in the slate (i.e., negated α-score). Most work on K-medians focuses on metric settings—our problem does not have such an interpretation—and little work has been done on non-metric settings (see, e.g., [1]) especially w.r.t. ordinal preferences. Facility location is another related problem, though the aim is usually to minimize the total cost of opening facilities and serving the nearest customers, with no constraints on the number of facilities. In our setting, the tradeoff between a positional score and the cost of alternatives is not well-defined unless the score is a surrogate for profit/cost.

Experiments on APA Dataset The American Psychological Association (APA) held a presidential election in 1980, where roughly 15,000 members expressed preferences for 5 candidates—5738 votes were full rankings. Members roughly divide into "academics" and "clinicians," who are on "uneasy terms," with classes of voters tending to favor one group of candidates over another (candidate groups $\{1, 3\}$ and $\{4, 5\}$ appeal to different voters, with candidate 2 somewhere in the middle) [9]. We apply our model to the full-ranking dataset with $K = 2$ and Borda scoring. We expect our model to favour "diverse" pairings (with academic-clinician pairings scoring highest). Indeed, this is what we obtain—the optimal recommendation set is $\{3, 4\}$ with $S_\beta = 18182$. In fact, the for highest scoring pairings are all diverse in this sense. Greedy outputs the diverse set $\{1, 5\}$ with score 17668, whereas selecting the top two candidates from the Borda or Kemeny rankings gives $\{1, 3\}$ with score 17352, an inferior (and non-diverse) pairing. The quality of the Borda/Kemeny approximations is even worse with more "dramatic" positional scoring (i.e., with scoring functions that exaggerate the score difference between different positions as discussed below).

Experiments on Sushi Dataset We experiment with a *sushi dataset* consisting of 10 varieties of sushi, and 5000 full preference orderings elicited across Japan [14]. In our budgeted (limited choice) setting, we might imagine a banquet in which only a small selection of sushi types can be provided to a large number of guests. Table 1 shows the approximation ratios of various algorithms for different slate sizes K, using an exponentially decreasing PSF $\alpha_{\text{exp}}(i) = 2^{m-i}$. CPLEX was used to solve IP (4) to determine optimal slates (computation times are shown in the table). We evaluate our greedy algorithm, random sets of size K (avg. over 20 instances for each K), and Borda and Kemeny (where we use the top K candidates as the recommendation set). We see that the Greedy algorithm always finds the optimal slate (and, in fact, does so for all $K \leq 9$), yet does so very quickly (under 1s.) relative to CPLEX optimization. Borda and Kemeny provide decent approximations, but are not generally optimal. Unsurprisingly, for large K (relative to $|A|$) random subsets do well, but perform poorly for small K. Results using Borda scoring are similar except that, unsurprisingly, random sets yield better approximations, since Borda count penalizes less for recommending lower-ranked alternatives than the exponential PSF.

In both the APA and sushi dataset, Borda and Kemeny rankings offer good approximations, though this is likely due, in part, to correlation effects: items that are highly preferred by an agent of one type are also reasonably preferred by agents of other types. This is in contrast to a situation (cf. Fig. 1) where one group's highly ranked candidate is strongly dispreferred by other groups.

K	Greedy	Borda	Kemeny	Random	CPLEX (sec.)
2	1.0	1.0	0.932	0.531	49.1
3	1.0	0.986	0.949	0.729	90.38
5	1.0	0.989	0.970	0.813	20.32
7	1.0	1.0	1.0	0.856	13.16

Table 1: Results on the sushi dataset with 10 alternatives and 5000 full rankings. Four algorithms are shown in the columns along with their approximation ratio for each K. CPLEX solution times are shown in the last column.

4 General Budgeted Social Choice

In the limited choice model, we assume the main bottleneck is the size of the recommendation set Φ. Once Φ is determined, voters are free to choose their favourite alternative. We can generalize the problem slightly by assigning costs to the alternatives and limiting the total cost of Φ (rather than its size). A more significant generalization involves also assuming some cost associated with each voter that benefits from an element in Φ. For example, a company that decides to manufacture different product configurations must pay certain fixed production costs for each configuration (e.g., capital expenditures); in addition, there are per-unit costs associated with producing each unit of the product (e.g., labour/material/transporation costs).[2]

For each alternative $a \in A$, let t_a be its *fixed cost* and u_a its *unit cost*. We assume a total budget B that cannot be exceeded by Φ. However, since unit costs vary across $a \in \Phi$, a decision maker cannot simply propose a recommendation set Φ: allowing agents to choose their most preferred alternative freely may result in exceeding the budget (e.g., if voters all choose expensive alternatives). Instead, the decision maker produces an *assignment* of alternatives to agents that maximizes social welfare.

Definition 8. *A* recommendation function $\Phi : N \to A$ *assigns agents to alternatives. Given PSF α and profile V, the* α-score *of Φ is:*

$$S_\alpha(\Phi, V) = \sum_{\ell \in N} \alpha_\ell(\Phi(\ell)) \; . \tag{8}$$

Let $\Phi(N) = \{a : \Phi^{-1}(a) \neq \emptyset\}$ be the set recommended alternatives. The cost *of Φ is:*

$$C(\Phi) = \sum_{a \in A} \mathbf{1}[a \in \Phi(N)] \cdot t_a + \sum_{\ell \in N} u_{\Phi(\ell)} \; . \tag{9}$$

The first component in the cost of Φ corresponds to the fixed costs of the recommended alternatives, and second reflects the total unit costs. We now define the general budgeted problem:

Definition 9. *Given alternatives A, profile V, PSF α and budget $B > 0$, the* budgeted social choice problem *is:*

$$\max_{\Phi} S_\alpha(\Phi, V) \quad \textit{subject to} \quad C(\Phi) \leq B. \tag{10}$$

We say that the problem is *infeasible* if every Φ has total cost exceeding B. As in the limited choice model, we define the problem using PSFs to measure utility and total social welfare as our optimization criterion; but other variants are possible. We mention a few interesting special cases:

- If we wish to leave some voters unassigned an alternative, we can model this using a *dummy item* d with $t_d = u_d = 0$. Voter preference for d can default to the bottom of each ordering or can reflect genuine preference for being unassigned. All such problems are feasible.

[2]The possibility of extending proportional representation schemes to making tradeoffs between representativeness and committee size is mentioned as an interesting possibility by Chamberlin and Courant [4].

- When $t_a = t$ (i.e., fixed charges are constant) and $u_a = 0$ for all $a \in A$, this corresponds to the limited choice model for $K = \lfloor B/t \rfloor$. Since unit costs are zero, the optimal Φ will always assign a voter to its preferred alternative, and a recommendation set of size B/t can be used. If unit costs are constant as well, $u_a = u$, similarly we have $K = \left\lfloor \frac{B-nu}{t} \right\rfloor$.
- When fixed costs vary, but unit costs $u = u_a$ are constant, we generalize the limited choice model slightly: because unit costs are identical, agents can still select their preferred alternative from a slate (of varying size) whose total fixed cost does not exceed $B - nu$.
- If every recommendation function Φ satisfies $C(\Phi) \leq B$ (e.g., if all charges are zero), we are in a *fully personalizable setting*, and each agent is assigned their their most preferred alternative.

Input: α, V, B, fixed costs t and unit costs u.

$\Phi \leftarrow \emptyset$ and $A^* \leftarrow \emptyset$
Let N_Φ denote $\{\ell : \Phi(\ell) \text{ is undefined}\}$
$\{$PHASE 1 : ADD ITEMS WITH BEST SWEET SPOT$\}$
loop
  for $a \in A \backslash A^*$ **do**
    $J \leftarrow \{\ell : a \succ_\ell r(\ell) \text{ and } u_a \geq u_{r(\ell)}\}$
    $N_a = \overline{N_\Phi} \cup J$
    $R_a = \left[\frac{\alpha_\ell(a)}{u_a}\right]_{\ell \in \overline{N_\Phi}} \cup \left[\frac{\alpha_\ell(a) - \alpha_\ell(\Phi(\ell))}{u_a - u_{\Phi(\ell)}}\right]_{\ell \in J}$
    $SR_a \leftarrow$ sort R_a to get $(\beta_1/\gamma_1, \ldots, \beta_{|R_a|}/\gamma_{|R_a|})$ $\{$If $\gamma_i = 0$ then the "ratio" gets put in front of sorted list. For another denominator $\gamma_j = 0$ we then compare whether $\beta_i > \beta_j$.$\}$
    reorder N_a to $[\ell_1^a, \ldots, \ell_{|N_a|}^a]$ so ℓ_i^a corresponds to β_i/γ_i
    Let r_a^* and i_a^* be the max and argmax over i of $\{\frac{\sum_{j=1}^{i} \beta_j}{t_a + \sum_{j=1}^{i} \gamma_j} : i \in |SR_a| \text{ and } t_a + \sum_{j=1}^{i} \gamma_j \leq B - C(\Phi)\}$ if $\emptyset$ then set to undefined.
  end for
  if $a^* \leftarrow \text{argmax}_{a \in A \backslash A^*} r_a^*$ is undefined **then**
    break $\{$*all* r_a^* is undefined—over budget$\}$
  else
    append a^* to A^*
    update Φ with $\{(\ell_i^{a^*}, a^*) : 1 \leq i \leq i_a^*\} \cup \{(\ell, a^*) : \ell \in N, a^* \succ_\ell \Phi(\ell) \text{ and } u_{a^*} \leq u_{\Phi(\ell)}\}$
  end if
end loop
$\{$PHASE 2: BACKTRACKING$\}$
while Φ incomplete **do**
  $a^* \leftarrow$ pop A^*
  remove $\{(\ell, a^*) : \ell \in N, \Phi(\ell) = a^*\}$ from Φ
  $\tilde{A} \leftarrow \{a \in A : t_a + \sum_{\ell \in \overline{N_\Phi}} u_a \leq B - C(\Phi)\}$
  if $\tilde{A} \neq \emptyset$ **then**
    $a^* \leftarrow \text{argmax}_{a \in \tilde{A}} \sum_{\ell \in \overline{N_\Phi}} \alpha_\ell(a)$
    update Φ with $\{(\ell, a^*) : \ell \in \overline{N_\Phi}\}$ and **break**
  end if
end while
return INFEASIBLE if $\Phi = \emptyset$, otherwise Φ

Fig. 2: The SweetSpotGreedy (SSG) algorithm.

We note that the general problem can be modified in other ways. For instance, we may ignore budget, and instead allow an explicit tradeoff between social welfare (voter happiness) and costs, and simply maximize total score less total cost of Φ. In this way, unit cost would not prevent assignment of some more preferred alternative to a voter if the voter's satisfaction outweighed the unit cost (once a fixed charge is incurred) or if it maximized surplus. This would better reflect a profit maximization motive in some settings (treating user satisfaction as a measure of willingness to pay). Our model as defined above is more appropriate in settings where users of a recommended alternative cannot be (directly) charged for its use (e.g., as in the case of certain public goods, corporate promotions or incentive programs, etc.).

Our general budgeted social choice problem is related to several problems arising in operations research. When fixed costs vary but unit costs are constant, the problem is similar to *budgeted maximum coverage* [16], given by a set E of weighted elements and a family of subsets of E with costs, with the goal of finding a covering with total cost under a budget that maximizes total weight of the covered elements. Our problem is slightly different: viewing voters as elements and alternatives as the cover set, we have a score for each element-alternative pair. Our problem is more closely related to the recently defined *generalized maximum coverage* problem [5], with a weight and cost for each cover set-element pair (in our model, the unit costs would be constant), and a cost for each cover set (i.e., fixed costs). Unlike budgeted social choice, coverage of all elements is not required. As discussed earlier, facility location is also similar to our budgeted setting, though it typically places no restrictions on budget (it is instead absorbed into the objective). Akin to unit costs in our model, [12] studies facility location when facility costs include the cost of customers being served (cost is assumed concave in number of customers).

We can formulate the general budgeted problem as an IP similar to IP (4) (with the same number of variables and constraints):

$$\begin{aligned} \max_{x_i, y_{\ell i}} \quad & (4) \\ \text{subject to} \quad & \left[\sum_{i=1}^{m} t_{a_i} x_i\right] + \left[\sum_{\ell \in N} \sum_{i=1}^{m} u_{a_i} y_{\ell i}\right] \leq B, \qquad (11) \\ & \text{and } (6), (7). \end{aligned}$$

An approximation algorithm for the general problem is complicated by the existence of unit costs. We may need to limit the assignment of expensive alternatives, despite "demand" from many voters. When unit costs are zero (or very low compared to fixed costs), the problem reduces to selecting a subset of alternatives as discussed above.

Still we develop a greedy heuristic algorithm called SweetSpotGreedy (or SSG). The main intuition behind our greedy heuristic is to successively "cover" or "satisfy" agents of a certain type by selecting their most preferred alternative. For a given $a \in A$, we sort voters based on their ranking of a and then compute the *bang-per-buck* ratio of assigning a to the first i voters—i.e., total score divided by total cost of assigning a to these i voters. We pick the index i^*_a that maximizes the bang-per-buck ratio r^*_a. This is the *sweet spot* since the marginal score improvement of assigning more a to additional voters doesn't justify the incremental cost of producing more of a. We then add to the recommendation function Φ that a^* with the greater ratio $r^*_{a^*}$ and assign it to the $i^*_{a^*}$ agents who prefer it most. We repeat this procedure after removing the previously assigned a, each time selecting a new a^* and recommending it to the voters that maximize its bang per buck. See Fig. 2 for further details. The first phase of the algorithm as described may not produce a feasible assignment Φ: the budget may be exhausted before all agents are assigned an alternative. A second backtracking phase produces a feasible solution by rolling back the most recent updates to Φ from Phase 1. Each time an alternative is rolled back, we try to find an $a \in A$ that can be assigned to all unassigned agents without depleting the budget. If after full backtracking this can't be achieved, the instance is infeasible (see Proposition 10).

SSG has running time $O(m^2 n \log n)$. The intuition behind our algorithm is similar in spirit to the $1 - \frac{1}{e} - o(1)$ approximation algorithm for generalized maximum coverage [5]. However, that algorithm is theoretical, requiring $O(m^2 n)$ calls to a fully polytime approximation scheme for the maximum density knapsack problem.

Proposition 10. SSG *returns INFEASIBLE iff the instance is infeasible.*

Proof. The if direction is obvious, since SSG always maintains feasibility of any solution Φ returned. If it returns INFEASIBLE, the backtracking phase must be entered and exited with $\Phi = \emptyset$. This implies $A^* = \emptyset$ since we have tried to roll back all additions to A^* only to discover there is no $a \in A$ with $t_a + n \cdot u_a \leq B$; that is, there is no single item assignable to all agents that doesn't exceed budget. This obviously implies infeasibility of the instance, since assigning the a minimizing $t_a + n \cdot u_a$ to all agents is the lowest cost Φ regardless of score. □

As discussed above, when unit costs are zero our problem reduces to selecting a subset $\Phi \subseteq A$ with total fixed cost less than B. When fixed costs are constant, this essentially reduces to the limited choice problem. In fact, SSG outputs the same recommendation function as that outputted by Greedy (converting the set to a function in the obvious way).

Proposition 11. *If* $u_a = 0$ *and* $t_a = 1$ *for all* $a \in A$ *then* SSG *outputs the same recommendation as* Greedy. *Hence, it has an approximation ratio* $1 - \frac{1}{e}$.

Proof. To see that SweetSpotGreedy reduces to Greedy notice that in the first iteration of Phase 1, Φ is empty, and because unit costs are zero, the sweet spot for any $a \in A$ is to recommend a to all $\ell \in N$. So Φ gets updated by assigning the alternative a_1^*, which maximizes the gain in total score, to all agents. On the next iteration, again because unit costs are zero, the sweet spot for any $a \in A - \{a_1^*\}$ is to recommend a to all agents ℓ that prefer it over a_1^*. Hence, Φ is updated by including the best alternative a_2^*. This observation holds in all subsequent iterations: the sweet spot for any unused alternative a is to recommend it to all agents who prefer a over existing elements of Φ. This is exactly what Greedy does, picking an alternative in each iteration (which is implicitly recommended to all agents that prefer it over existing alternatives in Φ) that greedily maximizes the gain in score. The $1 - \frac{1}{e}$ approximation ratio follows from Theorem 7. □

Experiments on Sushi Data We experimented with SweetSpotGreedy on the sushi dataset. In our first experiment, we randomly generate fixed costs while holding unit costs at zero. This corresponds to the special case discussed above that only slightly generalized the limited choice model. Integer fixed costs for the sushi varieties are chosen uniformly at random from $[20, 50)$, while the budget is set to 100. This means the recommendation set typically contains 2 to 5 items. We compared the performance of SSG against the optimal solution (computed using the IP above, solved using CPLEX) on 20 random instances (note that the preference profile is held fixed, corresponding to the data set). Both Borda scoring and the exponential PSF α_{exp} (see above) were tested and give similar results. With Borda, SSG is within 99% of the optimal recommendation function on average (it often attains the optimum, and is never worse than 94% of optimal). Its running times lie in the range $[1.91, 2.34]$ seconds (with a very simple Python implementation). Meanwhile, CPLEX has an average solution time of 114 seconds (the range is $[69s, 176s]$).

In a second experiment, we varied both fixed and unit costs with fixed costs substantially larger than unit costs. Specifically, integer unit costs were chosen uniformly at random from $[1, 4]$ and integer fixed costs from $[5000, 10000]$. We fixed the budget at 35000, which allows roughly 3 unique alternatives to be recommended. We again compare SSG to the optimal recommendation function on 20 random instances. Using Borda counts, the greedy algorithm gives recommendation functions that are, on average, within 98% of optimal, while taking 2–5s. to run. In contrast, CPLEX takes 458s. on average (range $[130s, 1058s]$) to produce an optimal solution. We achieve similar results using the exponential PSF, with greedy attaining average performance of 97% of optimal, and taking 3–6s. while CPLEX averages 321s. (range $[131s, 614s]$). These experiments show that SweetSpotGreedy has extremely strong performance, quickly finding excellent approximations to the optimal recommendation sets, when fixed costs are much larger than unit costs.

5 Conclusion

We have introduced a new class of *budgeted* social choice problems that spans the spectrum from genuine consensus (or "one-size-fits-all") recommendation typically studied in social choice to fully personalized decision-making. The key feature of our model—the fact that some customization to the preferences of distinct groups of users may be feasible where complete individuation is not—is characteristic of many real-world scenarios. Given a diverse array of user preferences, a decision maker must offer/produce/recommend a limited number of alternatives for the user population. This naturally leads to social welfare maximization goals whose solutions, crudely speaking, involve grouping/clustering agents with similar preferences and selecting one alternative for each group. Our model includes certain schemes for proportional representation as special cases, and indeed motivates the possible application for proportional schemes to ranking and recommendation. Such an objective often favours *diversity*, as opposed to *popularity*, of the chosen alternatives. This work can be viewed, for example, as justifying from social choice and decision-theoretic principles, that the top few web search results should be diversified so as to appeal to a wide range of user interests. We showed that the optimization induced by budgeted social choice is NP-hard; but we developed fast, intuitive greedy algorithms that have, in the case of the special case of limited choice, theoretical approximation guarantees. Critically, our greedy algorithms empirically provide excellent approximations on some real-world ordinal preference datasets.

Extensions of this work include the exploration of several variations of the budgeted model. For example, one might impose separate budgets for fixed and unit costs. If social welfare acts as a surrogate for the decision-maker's revenue/profit or return on investment, and the decision-maker has other investment options (e.g. a government considering public projects) one may wish to relax the budget constraints and instead maximize the return on investment per unit cost. Deeper connections to the proportional voting schemes is also being explored.

References

[1] M. R. Ackermann, J. Blömer, and C. Sohler. Clustering for metric and non-metric distance measures. *19th Symposium on Discrete Algorithms (SODA-08)*, pp.799–808, Philadelphia, 2008.

[2] K. J. Arrow. A difficulty in the concept of social welfare. *J. of Political Economy* 58:328–346, 1950.

[3] J. Bartholdi III, C. Tovey, and M. Trick. Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare* 6:157–165, 1989.

[4] J. R. Chamberlin and P. N. Courant. Representative deliberations and representative decisions: proportional representation and the Borda rule. *The American Political Science Review* 77(3):718–733, 1983.

[5] R. Cohen, L. Katzir. The generalized maximum coverage problem. *Inf. Process. Lett.*, 108:15–22, 2008.

[6] W. W. Cohen, R. E. Schapire, and Y. Singer. Learning to order things. *J. Artif. Intel. Research*, 10:243–270, 1998.

[7] V. Conitzer and T. Sandholm. Common voting rules as maximum likelihood estimators. *UAI-05*, pp.145–152, Edinburgh, 2005.

[8] J.-C. de Borda. Mémoire sur les élections au scrutin. *Histoire de l'Académie Royale des Sciences*, 1781.

[9] P. Diaconis. A generalization of spectral analysis with application to ranked data. *Annals of Statistics* 17(3):949–979, 1989.

[10] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. *World Wide Web Conference*, pp.613–622, New York, 2001.

[11] W. Gaertner. *A Primer in Social Choice Theory*. Oxford University Press, Oxford, 2006.

[12] M. Taghi Hajiaghayi, M. Mahdian, and V. S. Mirrokni. The facility location problem with general cost functions. *Networks*, 42(1):42–47, 2003.

[13] E. Hüllermeier, J. Fürnkranz, W. Cheng, and K. Brinker. Label ranking by learning pairwise preferences. *Artificial Intelligence*, 172(16-17):1897 – 1916, 2008.

[14] T. Kamishima, H. Kazawa, and S. Akaho. Supervised ordering: An empirical survey. *IEEE International Conference on Data Mining*, 673–676, 2005.

[15] J. Kemeny. Mathematics without numbers. *Daedalus*, 88(4):577–591, 1959.

[16] S. Khuller, A. Moss, and J. Naor. The budgeted maximum coverage problem. *Inf. Proc. Letters*, 70(1):39–45, 1999.

[17] T. Lu and C. Boutilier. The unavailable candidate model: A decision-theoretic view of social choice. *ACM Conference on Electronic Commerce 2010*, pp.263–274, Cambridge, MA.

[18] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions. *Mathematical Programming* 14(1):265–294, 1978.

[19] C. H. Papadimitriou, M. Schapira, and Y. Singer On the hardness of being truthful. *FOCS 2008* pp. 250–259.

[20] R. F. Potthoff and S. J. Brams. Proportional representation: broadening the options. *Journal of Theoretical Politics* 10:147–178, 1998.

[21] A. D. Procaccia, J. S. Rosenschein, and A. Zohar. On the complexity of achieving proportional representation. *Social Choice and Welfare* 30(3):353–362, 2008.

[22] P. Young. Optimal voting rules. *Journal of Economic Perspectives* 9:51–64, 1995.

Tyler Lu and Craig Boutilier

Dept. of Computer Science, University of Toronto, Toronto, CANADA

Email: `{tl,cebly}@cs.toronto.edu`