Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence

Finding a Collective Set of Items: From Proportional Multirepresentation to Group Recommendation

Piotr Skowron
University of Warsaw, Poland
p.skowron@mimuw.edu.pl

Piotr Faliszewski
AGH University, Poland
faliszew@agh.edu.pl

Jérôme Lang
Université Paris-Dauphine, France
lang@lamsade.dauphine.fr

Abstract

We consider the following problem: There is a set of items (e.g., movies) and a group of agents (e.g., passengers on a plane); each agent has some intrinsic utility for each of the items. Our goal is to pick a set of K items that maximize the total derived utility of all the agents (i.e., in our example we are to pick K movies that we put on the plane's entertainment system). However, the actual utility that an agent derives from a given item is only a fraction of its intrinsic one, and this fraction depends on how the agent ranks the item among the chosen, available, ones. We provide a formal specification of the model and provide concrete examples and settings where it is applicable. We show that the problem is hard in general, but we show a number of tractability results for its natural special cases.

1 Introduction

A number of real-world problems consist of selecting a set of items for a group of agents to jointly use. Such activities include, e.g., picking the movies to put on a plane's entertainment system, deciding which journals a university library should subscribe to, deciding what common facilities to build, or even choosing a parliament. Let us consider some common features of these examples.

First, there is a set of items[1] and a set of agents; each agent has some intrinsic utility for each of the items (e.g., this utility can be the level of appreciation for a movie, the average number of articles one reads from a given issue of a journal, expected benefit from building a particular facility, the feeling—measured in some way—of being represented by a particular politician).

Second, typically it is not possible to provide all the items to the agents and we can only pick some K of them, say (a plane's entertainment system fits only a handful of movies, the library has a limited budget, only several sites for the facilities are available, the parliament has a fixed size).

Third, the intrinsic utilities for items extend to the sets of items in such a way that the utility derived by an agent from a given item may depend on the *rank* of this item (from the agent's point of view) among the selected ones. Extreme examples include the case where each agent derives utility from his or her most preferred item only (e.g., an agent will watch his or her favorite movie only, will read/use the favorite journal/favorite facility only, will feel represented by the most appropriate politician only), from his or her least preferred item only (say, the agent worries that the family will force her to watch the worst available movie), or derives $1/K$ of the utility from each of the available items (e.g., the agent chooses the item—say, a movie—at random). However, in practice one should expect much more complicated schemes (e.g., an agent watches the top movie certainly, the second one probably, the third one perhaps, etc.; or, an agent is interested in having at least some T interesting journals in the library; an agent feels represented by some top T members of the parliament, etc.).

The goal of this paper is to formally define a model that captures all the above-described scenarios, to provide a set of examples where the model is applicable, and to provide an initial set of computational results for it in terms of efficient algorithms (exact or approximate) and computational hardness results (NP-hardness and inapproximability results).

Our work builds upon, generalizes, and extends quite a number of previously studied settings. We provide a deeper overview of this research throughout the paper. Here we mention the direct connection to the study of Chamberlin–Courant's multiwinner voting rule (see the original discussion of Chamberlin and Courant (1983) and, e.g., the papers of Procaccia et al. (2008), Betzler et al. (2013), Yu et al. (2013) and Skowron et al. (2013) for examples of computational results), to the study of budgeted social choice (Lu and Boutilier 2011; Oren and Lucier 2014), to the study of multiwinner approval voting rules (see, e.g., the overview of Kilgour (2010) and the papers of LeGrand et al. (2007) and Aziz et al. (2014)), and to several other settings, including, e.g., the Santa Claus problem (Bansal and Sviridenko 2006) and the problem of designing optimal picking sequences (Brams and Taylor 2000; Bouveret and Lang 2011; Kalinowski et al. 2012).

2 The Model

In this section we give a formal description of our model. However, before we move on to the mathematical details,

[1] We use the term 'item' in the most neutral possible way. Items may be candidates running for an election, or movies, or possible facilities, and so on.

let us explain and justify some high-level assumptions and choices that we have made.

First, we assume that the agents have separable preferences. This means that the *intrinsic utility* of an object does not depend on what other objects are selected. This is very different from, for example, the case of combinatorial auctions. However, in our model the *impact* of an object on the global utility of an agent does depend on its rank (according to that agent) among the selected items. This distinction between the intrinsic value of an item and its value distorted by its rank are also considered in several other research fields, especially decision theory ("rank-dependent utility theory") and multicriteria decision making, from which we borrow one of the main ingredients of our approach, *ordered weighted average (OWA) operators* (Yager 1988) (see the following formal model; the reader might want to consult the work of Kacprzyk et al. (2011) as well). OWAs were used recently in social choice in the context of rank-dependent scoring rules (Goldsmith et al. 2014): There the impact of the score obtained by a candidate from a vote depends on its rank in the list of scores it obtained from all votes.

Second, we assume that the agents' intrinsic utilities are provided explicitly in the input as numerical values, and that these values are comparable between agents (if one agent has twice as high a utility for some item than the other one, it means that this agent likes this item twice as much). Yet, we make no further assumptions about the nature of agents' utilities: they do not need to be normalized, they do not need to come from any particular range of values, etc. However, we consider several natural special cases of utility values (including approval-based utilities and Borda-based utilities; see the formal definitions later in this section).

Third, we take the *utilitarian* view and measure the social welfare of the agents as the sum of their perceived utilities. One could study other variants, such as the *egalitarian* variant, where the social welfare is measures as the utility of the worst-off agent. We leave this as future research (our preliminary attempts indicated that the egalitarian setting is even harder, computationally, than the utilitarian one).

The Formal Setting. Let $N = [n]$ be a set of n agents and let $A = \{a_1, \dots, a_m\}$ be a set of m items. The goal is to pick a size-K set W of items that, in some sense, is most satisfying for the agents. To this end, (1) for each agent $i \in N$ and for each item $a_j \in A$, we have an intrinsic utility u_{i,a_j}, $u_{i,a_j} \geq 0$, that agent i derives from a_j; (2) the utility that each agent derives from a set of K items is an ordered weighted average (Yager 1988) of this agent's intrinsic utilities for these items.

A weighted ordered average (OWA) over K numbers is a function defined through a vector $\alpha^{(K)} = \langle \alpha_1, \dots, \alpha_K \rangle$ of K (nonnegative) numbers[2] as follows: Let $\vec{x} = \langle x_1, \dots, x_K \rangle$ be a vector of K numbers and let $\vec{x}^{\downarrow} = \langle x_1^{\downarrow}, \dots, x_K^{\downarrow} \rangle$ be the nonincreasing rearrangement of $\vec{x}$, that is, $x_i^{\downarrow} = x_{\sigma(i)}$, where σ is any permutation of $\{1, \dots, K\}$ such that $x_{\sigma(1)} \geq x_{\sigma(2)} \geq \dots \geq x_{\sigma(K)}$. Then we set:

$$\mathrm{OWA}_{\alpha^{(K)}}(\vec{x}) = \textstyle\sum_{i=1}^{K} \alpha_i x_i^{\downarrow}$$

To make the notation lighter, we write $\alpha^{(K)}(x_1, \dots, x_K)$, instead of $\mathrm{OWA}_{\alpha^{(K)}}(x_1, \dots, x_K)$.

We provide a more detailed discussion of the OWA operators useful in our context later; we mention that they can be used, e.g., to express the arithmetic average (through the size-K vector $(\frac{1}{K}, \dots, \frac{1}{K})$), the maximum and minimum operators (through vectors $(1, 0, \dots, 0)$, and $(0, \dots, 0, 1)$, respectively) and the median operator (through the vector of all zeros, with a single one in the middle position).

We formalize our problem of computing "the most satisfying set of K items" as follows (also, see Example 1 below).

Definition 1. *In the* OWA-WINNER *problem we are given a set $N = [n]$ of agents, a set $A = \{a_1, \dots, a_m\}$ of items, a collection of agent's utilities $(u_{i,a_j})_{i \in [n], a_j \in A}$, a positive integer K ($K \leq m$), and a K-number OWA $\alpha^{(K)}$. The task is to compute a subset $W = \{w_1, \dots, w_K\}$ of A such that $u_{\mathrm{ut}}^{\alpha^{(K)}}(W) = \sum_{i=1}^{n} \alpha^{(K)}(u_{i,w_1}, \dots, u_{i,w_K})$ is maximal.*

For a family $(\alpha^{(K)})_{K=1}^{\infty}$ of OWAs, we write α-OWA-WINNER to denote the variant of the problem where, for a given solution size K, we use OWA $\alpha^{(K)}$. From now on we will not mention the size of the OWA vector explicitly and it will always be clear from context. We implicitly assume that OWAs in our families are polynomial-time computable.

Classes of Intrinsic Utilities. We often focus on variants of OWA-WINNER where agents' utilities are restricted.

Definition 2. *We say that the agents have approval-based utilities, if each agent's utilities come from the set $\{0, 1\}$.*

Definition 3. *Consider a setting with m items and let $u_{\max}$ denote the highest utility that some agent gives to an item. Let β and γ be two numbers in $[0, 1]$. We say that the agents have (β, γ)-non-finicky utilities if every agent has utility at least $\beta u_{\max}$ for at least γm items.*

Non-finicky utilities capture settings where each agent has relatively high utility for relatively many items. One of the most natural examples of non-finicky utilities are Borda-based utilities. Under Borda-based utilities, we assume that each agent ranks the items from the most desired one to the least desired one. Agent's utility of item a is the number of items that the agent prefers to a (so, if there are m items then an agent has utility $m - 1$ for its most preferred item, utility $m - 2$ for the second most preferred item, and so on). The order in which an agent ranks the items is the agent's *preference order*. Borda-based utilities are quite natural and, for example, were used in the original Chamberlin–Courant's rule and in several works on fair division (see, e.g., a paper of Brams and King (2005)).

Observation 1. *For every x, $0 \leq x \leq 1$, Borda-based utilities are $(x, 1 - x)$-non-finicky.*

There are other natural cases of non-finicky utilities. Consider agents that have approval-based utilities, where each agent approves at least a γ fraction of the items. These agents have $(1, \gamma)$-non-finicky utilities.

[2] The standard definition of OWAs assumes normalization, that is, $\sum_{i=1}^{K} \alpha_i = 1$. We do not make this assumption here, for the sake of convenience; note that whether OWA vectors are normalized or not is irrelevant to all notions and results of this paper.

Example 1. Consider six agents with the following Borda-based utilities over the items from the set $A = \{a_1, a_2, a_3, a_4, a_5, a_6\}$, expressed as preference orders:

$$\texttt{3 agents:}\ a_1 \succ a_2 \succ a_3 \succ a_5 \succ a_6 \succ a_4$$
$$\texttt{2 agents:}\ a_6 \succ a_1 \succ a_4 \succ a_3 \succ a_5 \succ a_2$$
$$\texttt{1 agent:}\ a_5 \succ a_4 \succ a_2 \succ a_3 \succ a_6 \succ a_1$$

We want to select $K = 3$ items and we use OWA $\alpha = (2, 1, 0)$. What is the score of $\{a_1, a_2, a_6\}$? The first three agents get utility $2 \times 5 + 4 = 14$ each, the next two get $2 \times 5 + 4 = 14$ each, and the last one gets $2 \times 3 + 1 = 7$. So, the score of $\{a_1, a_2, a_6\}$ is $42 + 28 + 7 = 77$. Indeed, this is the optimal set; the next best ones are $\{a_1, a_2, a_4\}$, $\{a_1, a_2, a_5\}$ and $\{a_1, a_5, a_6\}$, all with score 75. The rule defined by the OWA $\alpha' = (1, 1, 1)$, known as 3-Borda, would choose $\{a_1, a_2, a_3\}$ and Chamberlin–Courant's rule (in our terms, the rule defined by the OWA $\alpha'' = (1, 0, 0)$) would choose $\{a_1, a_5, a_6\}$.

A Dictionary of Useful OWA Families. Below we give a catalog of particularly interesting families of OWA operators (we take K to be the dimension of the vectors to which we apply a given OWA).

***k*-median OWA.** For each $k \in \{1, \dots, K\}$, $k\text{-med}^{(K)}$ is the OWA defined by the vector $(0, \dots, 0, 1, 0, \dots, 0)$, with the "1" on the k'th position; $k\text{-med}^{(K)}(\vec{x})$ is the k-th largest number among those in $\vec{x}$.

***k*-best OWA.** For each $k \in \{1, \dots, K\}$, $k\text{-best}^{(K)}$ OWA is the OWA defined by the vector $(1, \dots, 1, 0, \dots, 0)$ with k "1"s. That is, $k\text{-best}^{(K)}(\vec{x})$ is the sum of the top k values in $\vec{x}$ (with appropriate scaling, this means an arithmetic average of the top k numbers). $K\text{-best}_K^{(K)}$ is the sum of all the numbers in $\vec{x}$ (after scaling, their arithmetic average).

Arithmetic progression OWA. These OWAs are defined through vectors of the form $\text{aprog[a]}^{(K)} = \langle a + (K-1)b, a + (K-2)b, \dots, a\rangle$, where $a \geq 0$ and $b > 0$.

Geometric progression OWA. These OWAs are defined through vectors of the form $\text{gprog}[p]^{(K)} = \langle p^{K-1}, p^{K-2}, \dots, 1\rangle$, where $p > 1$.

Hurwicz OWA. For each λ, $0 \leq \lambda \leq 1$, Hurwicz OWA is defined through vector $(\lambda, 0, \dots, 0, 1-\lambda)$.

Other OWAs can be tailored for particular applications. We often, but not always, focus on OWAs defined through *non-increasing* vectors (this is natural since higher-ranked items should have more impact on the utility). Still, k-medians (except for 1-median) and Hurwicz OWAs (except for $\lambda = 1$) are natural OWAs that do not meet this criterion.

3 Applications of the Model

We believe that our model is very general. To substantiate this claim, in this section we provide four, quite different, scenarios where it is applicable.

Generalizing Voting Rules. Our research started as an attempt to generalize the rule of Chamberlin and Courant (1983) for electing sets of representatives. For this rule, the voters (the agents) have Borda-based utilities over a set of candidates and we wish to elect a K-member committee (e.g., a parliament), such that each voter is represented by one member of the committee. If we select K candidates, then a voter is "represented" by the selected candidate that she ranks highest among the chosen ones. Thus, winner determination under Chamberlin–Courant's voting rule boils down to solving 1-best-OWA-WINNER for the case of Borda-based utilities. On the other hand, solving K-best-OWA-WINNER for Borda-based utilities is equivalent to finding winners under K-Borda, the rule that picks K candidates with the highest Borda scores (see the work of Elkind et al. (2014) for a classification of multiwinner voting rules, including, e.g., K-Borda and Chamberlin–Courant's rule).

Our model extends one more appealing voting rule, known as Proportional Approval Voting (PAV; see the works of Kilgour (2010), for a review, and of Aziz et al. (2014), for computational results). Winner determination under PAV is equivalent to solving α-OWA-WINNER for the OWA vector $\alpha = \langle 1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{m}\rangle$, with approval-based utilities.

Malfunctioning Items or Unavailable Candidates. Consider a setting where we pick the items off-line, but on-line it may turn out that some of them are unavailable (for example, we pick a set of journals the library subscribes to, but when an agent goes to a library, a particular journal could already be borrowed to someone else; see the work of Lu and Boutilier (2010) for other examples of social choice with possibly unavailable candidates). We assume that each item is available with the same, given, probability p (i.i.d.). The utility an agent gets from a set of selected items W is the expected value of the best available object. The probability that the i'th item is available while the preceding $i-1$ items are not, is proportional to $p(1-p)^{i-1}$. So, to model the problem of selecting items in this case, we should use the geometric progression OWA with initial value p and coefficient $1-p$.

Uncertainty Regarding How Many Items a User Enjoys. There may be some uncertainty about the number of items a user would enjoy (e.g., on a plane, it is uncertain how many movies a passenger would watch; one might fall asleep or might only watch those movies that are good enough). We give two possible models for the choice of the OWA vectors:

1. The probability that an agent enjoys i items, for $0 \leq i \leq K$, is uniformly distributed, i.e., an agent would enjoy exactly his or her first i items in W with probability $\frac{1}{K+1}$. So, the agent enjoys the i'th item if she enjoys at least i items, which occurs with probability $\frac{K-i+1}{K+1}$; we should use OWA vector defined by $\alpha_i = K - i + 1$ (we disregard the normalizing constant), i.e., an arithmetic progression.
2. We assume that the values given by each user to each item are distributed uniformly, i.i.d., on $[0, 1]$ and that each user uses only the items that have a value at least θ, where θ is a fixed (user-independent) threshold. Therefore, a user enjoys the item in W ranked in position i if she values at least i items at least θ, which occurs with probability $\sum_{j=i}^{K} \binom{K}{i}(1-\theta)^i\theta^{K-i}$, thus leading to the OWA vector defined by $\alpha_i = \sum_{j=i}^{K} \binom{K}{i}(1-\theta)^i\theta^{K-i}$.

Ignorance About Which Item Will Be Assigned to a User. We now assume that a matching mechanism will be used after selecting the K items. The matching mechanism is not specified; it might also be randomized. If the agents have a *complete ignorance about the mechanism used*, then it makes sense to use known criteria for decision-making under complete uncertainty:

1. The *Wald* criterion assumes that agents are extremely risk-averse, and corresponds to $\alpha = K\text{-med}^{(K)}$. The agents consider their worst possible items.
2. The *Hurwicz* criterion is a linear combination between the worst and the best outcomes, and corresponds to $\alpha = (\lambda, 0, \dots, 0, 1-\lambda)$ for some fixed $\lambda \in (0,1)$.

If the agents know that they are guaranteed to get one of their best i items, then the Wald and Hurwicz criteria lead, respectively, to the OWAs $\alpha = i\text{-med}^{(K)}$ and $\alpha = (\lambda, 0, \dots, 0, 1-\lambda, 0, \dots, 0)$, with $1-\lambda$ in position i. If the agents know that the mechanism gives them one of their top i items, each with the same probability, then we should use i-best OWA. More generally, the matching mechanism may assign items to agents with a probability that decreases when the rank increases.

4 Computational Results

We study the complexity and the approximability of our problem, focusing in its most natural special cases. Missing proofs are available in our technical report (Skowron, Faliszewski, and Lang 2014). Our results are summarized in Table 1.

Theorem 1. *The complexity and (in)approximability of* OWA-WINNER *is as stated in Table 1.*

The text below focuses on those contributions that we found either most essential or most interesting.

Nonincreasing OWAs. In general, OWA-WINNER is NP-hard (this follows, e.g., from the work of Procaccia et al. (2008) on Chamberlin–Courant's rule). We show that this holds for each nontrivial, nonincreasing OWA operator.

Theorem 2. *Fix an OWA family α, such that for every K, $\alpha^{(K)}$ is nonincreasing and nonconstant; α-*OWA-WINNER *is* NP*-hard for approval-based utilities.*

This theorem has some quite interesting consequences. For example, if K is the number of items that we are to pick, it shows that $(K-1)$-best-OWA-WINNER is NP-hard, even though K-best-OWA-WINNER is in P (indeed, through a fairly technical proof, these results hold even for Borda-based utilities). Nonetheless, for nonincreasing OWAs a simple greedy algorithm achieves $(1-1/e)$ approximation ratio. This follows by applying the famous result of Nemhauser et al. (1978) for nondecreasing submodular set functions to the case of u_{ut}^{α} (recall Definition 1). If A is some set and u is a function $u\colon 2^A \to \mathbb{R}_+$, then we say that: (1) u is submodular if for each W and W', $W \subseteq W' \subseteq A$, and each $a \in A \setminus W'$ it holds that $u(W \cup a) - u(W) \geq u(W' \cup a) - u(W')$, and (2) u is nondecreasing if for each $W \subseteq A$ and each $a \in A$ it holds that $u(W \cup \{a\}) \geq u(W)$.

Table 1: Our computational results for OWA-WINNER. For each OWA family we provide four entries: In the first row (for a given OWA family) we give its worst-case complexity and in the second row we give our best approximation result. In the first column the results are for approval-based utilities, and in the second one they are for both Borda-based utilities and $(1,\gamma)$-non-finicky utilities (NP-hardness results apply to Borda, approximation results apply to both). The result marked with ♠ shows that an approximation result for OWA-WINNER would give the same approximation guarantee for the DENSEST-K-SUBGRAPH problem (Bhaskara et al. 2012; Raghavendra and Steurer 2010); there is evidence suggesting this problem is hard to approximate. The result marked with ♡ is derived from inapproximability of the MAXIMUM EDGE BICLIQUE PROBLEM (Feige and Kogan 2004); under mild assumptions it shows inapproximability up to any constant factor (and even certain factors depending on n).

OWA family	general utilities	Borda/ $(1,\gamma)$–non-finicky utilities
k-median (k fixed)	NP-hard inapprox. result♠	NP-hard (Borda) PTAS
K-median	NP-hard inapprox. result♡	NP-hard (Borda) ?
k-best (k fixed)	NP-hard $(1-\frac{1}{e})$-approx.	NP-hard (Borda) PTAS
$(K-1)$-best	NP-hard PTAS	NP-hard (Borda) PTAS
K-best	P	P
arithmetic progres.	NP-hard $(1-\frac{1}{e})$-approx.	? $(1-\frac{1}{e})$-approx.
geometric progres.	NP-hard $(1-\frac{1}{e})$-approx.	? PTAS
Hurwicz[λ]	NP-hard $\lambda(1-\frac{1}{e})$-approx.	? $\lambda(1-\epsilon)$-approx. for each $\epsilon > 0$

Lemma 3. *Let I be an instance of* OWA-WINNER *with a nonincreasing OWA α. The function u_{ut}^{α} is submodular and nondecreasing.*

Proof. Let I be an instance of OWA-WINNER with agent set $N = [n]$, item set $A = \{a_1, \dots, a_m\}$, desired solution size K, and OWA $\alpha = \langle \alpha_1, \dots, \alpha_K \rangle$. For each agent $i \in N$ and each item $a_j \in A$, u_{i,a_j} is a nonnegative utility that i derives from a_j.

Since all the utilities and all the entries of the OWA vector are nonnegative, we note that u_{ut}^{α} is nondecreasing. To show submodularity, we decompose u_{ut}^{α} as follows: $u_{\text{ut}}^{\alpha}(W) = \left(\sum_{\ell=1}^{K-1}(\alpha_\ell - \alpha_{\ell+1}) u_{\text{ut}}^{\ell\text{-best-OWA}}(W)\right) + \alpha_K u_{\text{ut}}^{K\text{-best-OWA}}(W)$. For each $W \subseteq A$, $i \in N$ and $\ell \in [m]$, let $\mathrm{Top}(W, i, \ell)$ be the set of those ℓ items from W whose utility, from the point of view of agent i, is highest (we break ties in an arbitrary way). Since nonnegative linear combinations of submodular functions are submodular, it suffices to prove that for each $i \in$

N and each $\ell \in [m]$, function $u_i^\ell(W) = \sum_{w \in \mathrm{Top}(W,i,\ell)} u_{i,w}$ is submodular.

To show submodularity of u_i^ℓ, consider two sets, W and W', $W \subseteq W' \subseteq A$, and some $a \in A \setminus W'$. We claim that:

$$u_i^\ell(W \cup \{a\}) - u_i^\ell(W) \geq u_i^\ell(W' \cup \{a\}) - u_i^\ell(W'). \quad (1)$$

Let u_W and $u_{W'}$ denote the utilities that the i-th agent has for the ℓ-th best items from W and W', respectively (or 0 if a given set has fewer than ℓ elements). Of course, $u_{W'} \geq u_W$. Let u_a denote i-th agent's utility for a. We consider two cases. If $u_a \leq u_W$, then both sides of (1) have value 0. Otherwise $u_i^\ell(W' \cup \{a\}) - u_i^\ell(W') = \max(u_a - u_{W'}, 0)$ and $u_i^\ell(W \cup \{a\}) - u_i^\ell(W) = u_a - u_W$, which proves (1) and completes the proof. □

Algorithm 1. Select K items greedily, each time choosing an item that increases the utility of the agents most (maximizes the function u_{ut}^α).

The next result follows by applying the result of Nemhauser et al. (1978) to function u_{ut}^α and Algorithm 1.

Theorem 4. *For a nonincreasing OWA α, Algorithm 1 is a polynomial time $(1 - 1/e)$-approximation algorithm for the problem of finding the utilitarian set of K winners.*

Algorithm 1 can be seen as a new multiwinner voting rule on its own. Indeed, many popular voting rules are defined as iterative (greedy) algorithms; such rules are not only polynomially solvable, but sometimes easier to understand for the society than those defined declaratively. The view of approximation algorithms as voting rules on their own was initiated by Caragiannis et al. (2012; 2010) and was continued for the case of multiwinner rules by Skowron et al. (2013) and Elkind et al. (2014), where this last paper focused on the properties of multiwinner rules. Here, we give another interesting observation. Using Theorem 4, we found out that Sequential Proportional Approval Voting, developed by the Danish polymath Thorvald N. Thiele, and used for a short period in Sweden during early 1900's (see the work of Brams and Kilgour (2010)) is simply a greedy approximation algorithm for PAV (see the discussion in Section 3). This observation gives another evidence that approximation algorithms for computationally hard voting rules, can, indeed, be viewed as full-fledged voting rules.

Going back to Theorem 4, we should wonder if a $(1 - \frac{1}{e})$-approximation algorithm is a good result. Irrespective if one views it as sufficient or not, this is the best possible approximation ratio a polynomial-time algorithm can have for OWA-WINNER with a nonincreasing OWA and unrestricted utilities. The reason is that 1-best-OWA-Winner with approval-based utilities is, in essence, another name for the MAXCOVER problem, and if P $\neq$ NP, then $(1 - \frac{1}{e})$ is the approximation upper bound for MAXCOVER (Feige 1998). Better approximation bounds are possible only for very particular OWA vectors. For example, we show a polynomial-time approximation scheme (PTAS) for $(K-1)$-best-OWA-WINNER (where K is the size of the set that we pick) that works irrespective of the nature of agent utilities.

OWAs That Are Not Nonincreasing. The assumption of the OWAs being nonincreasing is important. In the full version of the paper (Skowron, Faliszewski, and Lang 2014) we provide evidence that without it, typically our problem cannot be approximated up to any constant factor (under usual complexity-theoretic assumptions). In particular, we show that for such OWAs function $u_{\mathrm{ut}}^{\alpha^{(K)}}(W)$ is not submodular.

Algorithm 2: For nonincreasing OWAs where at most first ℓ entries are nonzero, for (β, γ)-non-finicky utilities.

Notation:
$\Phi \leftarrow$ a map defining the number of free slots per agent.
Initially for each agent i we have $\Phi[i] = \ell$.

```
x ← γm;
W ← ∅;
for i ← 1 to K do
    a ← argmax_{a∈A\W} ‖{j | Φ(j) > 0 ∧ ‖{b : u_{j,b} > u_{j,a}}‖ < x}‖;
    foreach j ∈ {j | Φ(j) > 0} do
        if ‖{b : u_{j,b} > u_{j,a}}‖ < x then
            Φ[j] ← Φ[j] − 1;
    W ← W ∪ {a};
return W
```

Non-Finicky Utilities. One of the greatest sources of hardness of OWA-WINNER, that we rely on in our proofs, is that the agents may have very high utilities for some very small subsets of items, and very low utilities for the remaining ones (consider, e.g., approval-based utilities where each agent approves of relatively few items). In such cases, intuitively, either we find a perfect solution or some of the agents have to be very badly off. On the other hand, for Borda-based utilities when some agent does not get his top items, it is still possible to provide the agent with not-much-worse ones; the utilities decrease linearly. Indeed, Skowron et al. (2013) used this observation to give a PTAS for the Chamberlin–Courant's rule. Here we give a strong generalization of their result that applies to appropriate non-finicky utilities and OWA families including k-median, k-best, and geometric progression OWAs. Due to restricted space, we provide the most interesting proof only.

We focus on the case of OWA vectors where only some constant number ℓ of top positions are nonzero, and on (β, γ)-non-finicky utilities ($\beta, \gamma \in [0, 1]$). In this case, Algorithm 2 (a generalization of an algorithm of Skowron et al. (2013)) achieves a good approximation ratio. The idea behind the algorithm is as follows: To pick K items, it proceeds in K iterations and in each iteration it introduces one new item into the winner set. For each agent it considers the top $x = \gamma m$ items with the highest utilities and in a given iteration it picks an item a that maximizes the number of agents that (1) rank a among items with the highest x utilities, and (2) still have "free slots" (an agent has a free slot if among the so-far-selected winners, fewer than ℓ have utilities among the x highest ones for this agent).

Theorem 5. *Fix a positive integer ℓ and let α be a nonincreasing OWA where at most first ℓ entries are nonzero. If the agents have (β, γ)-non-finicky utilities, with $\gamma m \geq \ell$, then Algorithm 2 is a polynomial-time $\beta(1 - \exp(-\frac{\gamma K}{\ell}))$-approximation algorithm for α-OWA-WINNER.*

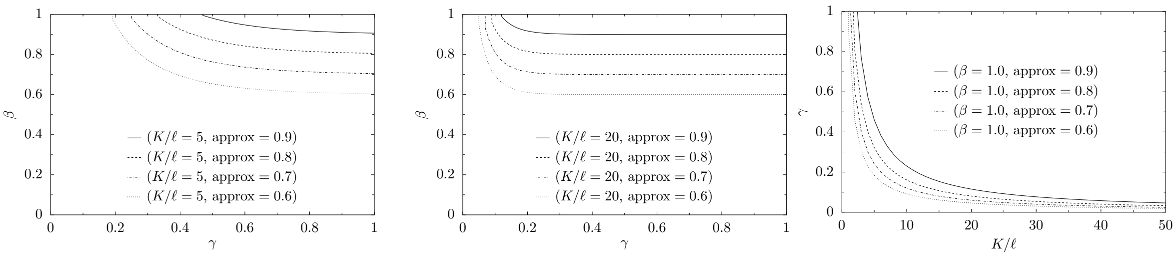

Figure 1: The approximation ratios of Algorithm 2 for a nonincreasing OWA with at most ℓ top coefficients greater than zero, for (β, γ)-non-finicky utilities. The lines in Figures (a) and (b) depict the relations between the parameters β and γ that, for a given fixed ratio $\frac{K}{\ell}$ lead to the same approximation bound. The lines in Figure (c) depict the relations between the parameter γ and the ratio K/ℓ that, for a given fixed value of the parameter β lead to the same approximation bound.

Proof. Consider an instance I of α-OWA-WINNER, with n agents, m items, and where we seek a winner set of size K. Let $x = \gamma m$. We use an OWA where an agent's total utility from a winner set W depends on this agent's utilities for his or her top ℓ items from W. Thus, we introduce the notion of each agent's free slots. Initially, each agent has ℓ free slots. Whenever an agent j has a free slot and the algorithm selects an item a such that for j, a is among x items with highest utilities, we say that a starts occupying one free slot of j. After such an item is selected, j has one free slot less.

Let n_i denote the total number of free slots of all the agents after the i-th iteration of the algorithm; $n_0 = \ell n$. We show by induction that $n_i \leq \ell n \left(1 - \frac{x}{\ell m}\right)^i$. Indeed, the inequality is true for $i = 0$. Let us assume that it is true for some i: $n_i \leq \ell n \left(1 - \frac{x}{\ell m}\right)^i$. Let F_i denote the set of agents that have free slots after iteration i. There are at least $\frac{n_i}{\ell}$ such agents. For $j \in F_i$, let $S(j)$ be the number of j's top-x items that were not included in the solution yet. If $j \in F_i$ has s free slots, then $S(j) = (x - \ell + s)$. Thus we have that $\sum_{j \in F_i} S(j) \geq n_i + (x - \ell)\frac{n_i}{\ell} = \frac{n_i x}{\ell}$. By the pigeonhole principle, there exists an item that is among top-x items for at least $\frac{n_i x}{\ell m}$ agents from F_i. Thus, after the $(i+1)$-th iteration of the algorithm, the total number of free slots is at most:

$$n_{i+1} \leq n_i - \frac{n_i x}{\ell m} = n_i \left(1 - \frac{x}{\ell m}\right) \leq \ell n \left(1 - \frac{x}{\ell m}\right)^{(i+1)}.$$

The number of free slots after the last iteration is at most:

$$n_K \leq \ell n \left(1 - \frac{x}{\ell m}\right)^K = \ell n \left(1 - \frac{\gamma}{\ell}\right)^K \leq \ell n \exp\left(-\frac{\gamma K}{\ell}\right).$$

Thus, the number of occupied slots is at least $\ell n - \ell n \exp(-\frac{\gamma K}{\ell})$. Every item that occupies an agent's slot has utility for this agent at least $\beta u_{\max}$, where $u_{\max}$ is the maximal utility that any of the agents assigns to an item.

It remains to assess the OWA coefficients for the utilities of the items in the solution. If for some agent i the utility of an item a, $u_{i,a}$, is taken with coefficient α_p ($p > 1$), then in the solution there must be an item a' such that $u_{i,a'} \geq u_{i,a}$ and $u_{i,a'}$ is taken with coefficient α_{p-1}. So there must exist at least $\frac{1}{\ell}(\ell n - \ell n \exp(-\frac{\gamma K}{\ell}))$ occurrences of the items whose utilities are taken with coefficient α_1. By repeating this reasoning for the remaining occurrences of the items from the solution, since α is nonincreasing, we get that the total utility of the agents is at least $\beta u_{\max}(\ell n - \ell n \exp(-\frac{\gamma K}{\ell}))\frac{1}{\ell}\sum_{i=1}^{\ell} \alpha_i = \beta u_{\max} n(1 - \exp(-\frac{\gamma K}{\ell}))\sum_{i=1}^{\ell} \alpha_i$. Since no solution has utility higher than $n u_{\max} \sum_{i=1}^{\ell} \alpha_i$, we get our approximation ratio. □

Approximation ratio of Algorithm 2 is particularly good when K is large compared to ℓ. This, indeed, is the most interesting case because for small K we can find optimal solutions by brute-force search (combining these two approaches leads to a PTAS; see Theorem 6 below). Nevertheless, Algorithm 2 often gives a satisfactory approximation guarantees by itself. Figure 1 depicts the classes of non-finicky utilities for which, for a fixed ratio K/ℓ, Algorithm 2 guarantees appropriate approximation ratios: Parts (a) and (b) of the figure show the relation that β and γ have to satisfy to obtain a particular approximation ratio, for a given value $\frac{K}{\ell}$. Part (c) shows the relation between the value of γ and the ratio $\frac{K}{\ell}$ that has to be satisfied for Algorithm 2 to achieve a particular approximation ratio under $(1, \gamma)$-non-finicky utilities.

Interestingly, Theorem 5 can be generalized to the case of OWAs that are not nonincreasing (achieving a slightly weaker approximation ratio). Combining this result with a brute-force search for small values of K, we obtain a PTAS.

Theorem 6. *Fix a value ℓ and let α be a family of OWAs that have nonzero values on top ℓ positions only. There is a PTAS for α-OWA-WINNER for the case of (i) Borda-based utilities, and (ii) $(1, \gamma)$-non-finicky utilities (for constant γ).*

5 Summary

We proposed a new model for the problem of selecting a number of items used jointly by a group of agents. We argued that the model is practical and we showed a number of settings where it is directly relevant. In particular, it generalizes several known multiwinner voting rules. We investigated tractability of our problems. We have provided general hardness results and a number of approximation algorithms for natural special cases of the problem.

Acknowledgements. The authors were supported by NCN grant 2013/09/N/ST6/03661 and by AGH University grant 11.11.230.124 (statutory research).

References

Aziz, H.; Gaspers, S.; Gudmundsson, J.; Mackenzie, S.; Mattei, N.; and Walsh, T. 2014. Computational aspects of multi-winner approval voting. In *8th Multidisciplinary Workshop on Advances in Preference Handling (MPREF-14)*.

Bansal, N., and Sviridenko, M. 2006. The Santa Claus problem. In *Proceedings of STOC-2006*, 31–40. ACM Press.

Betzler, N.; Slinko, A.; and Uhlmann, J. 2013. On the computation of fully proportional representation. *Journal of Artificial Intelligence Research* 47(1):475–519.

Bhaskara, A.; Charikar, M.; Vijayaraghavan, A.; Guruswami, V.; and Zhou, Y. 2012. Polynomial integrality gaps for strong sdp relaxations of densest k-subgraph. In *Proceedings of SODA-2012*, 388–405.

Bouveret, S., and Lang, J. 2011. A general elicitation-free protocol for allocating indivisible goods. In *Proceedings of IJCAI-2011*, 73–78.

Brams, S. J., and Kilgour, D. M. 2010. Satisfaction approval voting. Mpra paper, University Library of Munich, Germany.

Brams, S., and King, D. 2005. Efficient fair division help the worst off or avoid envy? *Rationality and Society* 17(4):387–421.

Brams, S., and Taylor, A. 2000. *The Win-Win solution: Guaranteeing Fair Shares to Everybody*. WW Norton & Company.

Caragiannis, I.; Kaklamanis, C.; Karanikolas, N.; and Procaccia, A. 2010. Socially desirable approximations for Dodgson's voting rule. In *Proceedings of EC-2010*, 253–262.

Caragiannis, I.; Covey, J.; Feldman, M.; Homan, C.; Kaklamanis, C.; Karanikolas, N.; Procaccia, A.; and Rosenschein, J. 2012. On the approximability of Dodgson and Young elections. *Artificial Intelligence* 187:31–51.

Chamberlin, B., and Courant, P. 1983. Representative deliberations and representative decisions: Proportional representation and the borda rule. *American Political Science Review* 77(3):718–733.

Elkind, E.; Faliszewski, P.; and Slinko, P. S. A. 2014. Properties of multiwinner voting rules. In *Proceedings of AAMAS-2014*, 53–60.

Feige, U., and Kogan, S. 2004. Hardness of approximation of the balanced complete bipartite subgraph problem. Technical report.

Feige, U. 1998. A threshold of ln n for approximating set cover. *Journal of the ACM* 45(4):634–652.

Goldsmith, J.; Lang, J.; Mattei, N.; and Perny, P. 2014. Voting with rank dependent scoring rules. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, 698–704.

Kacprzyk, J.; Nurmi, H.; and Zadrozny, S. 2011. The role of the OWA operators as a unification tool for the representation of collective choice sets. In *Recent Developments in the Ordered Weighted Averaging Operators*. 149–166.

Kalinowski, T.; Narodytska, N.; Walsh, T.; and Xia, L. 2012. Elicitation-free protocols for allocating indivisble goods. In *Proceedings of COMSOC-2012*.

Kilgour, D. 2010. Approval balloting for multi-winner elections. In Laslier, J., and Sanver, R., eds., *Handbook on Approval Voting*. Springer. 105–124.

LeGrand, R.; Markakis, E.; and Mehta, A. 2007. Some results on approximating the minimax solution in approval voting. In *Proceedings of AAMAS-2007*, 1193–1195.

Lu, T., and Boutilier, C. 2010. The unavailable candidate model: a decision-theoretic view of social choice. In *Proceedings of ACM-EC-2010*, 263–274.

Lu, T., and Boutilier, C. 2011. Budgeted social choice: From consensus to personalized decision making. In *Proceedings of IJCAI-2011*, 280–286. See also the version that appeared in COMSOC-10.

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

Oren, J., and Lucier, B. 2014. Online (budgeted) social choice. In *Proceedings of AAAI-2014*, 1456–1462. AAAI Press.

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

Raghavendra, P., and Steurer, D. 2010. Graph expansion and the unique games conjecture. In *Proceedings of STOC-2010*, 755–764.

Skowron, P.; Faliszewski, P.; and Lang, J. 2014. Finding a collective set of items: From proportional multirepresentation to group recommendation. Technical Report arXiv:1402.3044 [cs.MA].

Skowron, P.; Faliszewski, P.; and Slinko, A. 2013. Fully proportional representation as resource allocation: Approximability results. In *Proceedings of IJCAI-2013*.

Yager, R. 1988. On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Transactions on Systems, Man and Cybernetics* 18(1):183–190.

Yu, L.; Chan, H.; and Elkind, E. 2013. Multiwinner elections under preferences that are single-peaked on a tree. In *Proceedings of IJCAI-2013*, 425–431.