Artificial Intelligence 241 (2016) 191–216

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Finding a collective set of items: From proportional multirepresentation to group recommendation ☆

Piotr Skowron [a,*], Piotr Faliszewski [b], Jérôme Lang [c]

[a] University of Oxford, Oxford, UK
[b] AGH University of Science Technology, Krakow, Poland
[c] Université Paris-Dauphine, Paris, France

ARTICLE INFO

Article history:
Received 29 October 2015
Received in revised form 3 June 2016
Accepted 16 September 2016
Available online 22 September 2016

Keywords:
Proportional representation
Ordered weighted average
Chamberlin–Courant rule
Computational complexity
Approximation
Elections
Voting

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. Examples of such activities include picking a set of 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 voting for a parliament (or other assembly of representatives). 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

☆ The preliminary version of this paper was presented at AAAI-2015.

* Corresponding author.
E-mail addresses: piotr.skowron@cs.ox.ac.uk (P. Skowron), faliszew@agh.edu.pl (P. Faliszewski), lang@lamsade.dauphine.fr (J. Lang).

[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.

http://dx.doi.org/10.1016/j.artint.2016.09.003
0004-3702/

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 him or 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, provide a set of examples where the model is applicable, and 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 settings that have already been studied in the literature. We provide a deeper overview of this research in Section 8 and here we only mention the two most directly related lines of work. First, our model where the agents derive utility from their most preferred item among the selected ones directly corresponds to winner determination under the Chamberlin–Courant's voting rule [18,50,7] (it is also very deeply connected to the model of budgeted social choice [41,49,42]) and is in a certain formal sense a variant of the facility location problem. Second, the case where for each item each agent derives the same fraction of the utility is, in essence, the same as K-winner range-voting (or K-winner Borda [21]); that agents enjoy equally the items they get is also a key assumption in the Santa Claus problem [6], and in the problem of designing optimal picking sequences [14,10,35].

The paper is organized as follows. First, in Section 2 we discuss several important modeling choices and provide the formal description of our model. Then, in Section 3, we discuss the applicability of the model in various scenarios. Specifically, we show a number of examples that lead to particular parameter values of our model. We give an overview of our results in Section 4 and then, in Sections 5, 6, and 7, we present these results formally. In Section 5 we present results regarding the complexity of computing exact solutions for our model. In the next two sections we discuss the issue of computing approximate solutions. First without putting restrictions on agents' utilities (Section 6) and, then, for what we call non-finicky utilities (Section 7). Intuitively put, under non-finicky utilities the agents are required to give relatively high utility values to a relatively large fraction of the items. We believe that the notion of non-finicky utilities is one of the important contributions of this paper. We discuss related work in Section 8 and conclude in Section 9.

2. The model

In this section we give a formal description of our model. However, before we move on to the mathematical details, 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 is also considered in several other research fields, especially in decision theory (where it is known as "rank-dependent utility theory") and in multicriteria decision making, from which we borrow one of the main ingredients of our approach, the *ordered weighted average (OWA) operators* [58] (for technical details see the work of Kacprzyk et al. [34]). OWAs were recently used in social choice in several contexts [31,3,23]; we discuss these works in detain in Section 8.

Second, throughout the paper we navigate between two views of the agents' intrinsic utilities:

1. Generally, we assume that the utilities are provided explicitly in the input as numerical values, and that these values are comparable between agents. 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. Indeed, it is possible that some agent has very strong preferences regarding the items, modeled through high, diverse utility values, whereas some other agent does not care much about the selection process and has low utility values only.
2. In some parts of the paper (which will always be clearly identified), we assume that utilities are heavily constrained and are derived from non-numerical information, such as approval ballots specifying which items an agent approves (leading to approval-based utilities), or rankings over alternatives, from which utilities are derived using an agent-independent scoring vector (typically, a Borda-like vector).

Formally, the latter view is a special case of the former, but we believe that it is worthwhile to consider it separately. Indeed, many multiwinner voting rules (such as the Chamberlin–Courant [18] rule or the Proportional Approval Voting rule [37]) fit the second view far more naturally, whereas for other applications the former view is more natural.

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 measured as the utility of the worst-off agent. We leave this as possible future research (our preliminary attempts indicated that the egalitarian setting is computationally even harder than the utilitarian one). Very recently, Elkind and Ismaïli [23] used OWA operators to define variants of the Chamberlin–Courant rule that lay between the utilitarian and egalitarian variants, while Amanatidis et al. [3]