Artificial Intelligence 263 (2018) 74–106

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Multi-attribute proportional representation ☆

Jérôme Lang [a], Piotr Skowron [b]

[a] Université Paris-Dauphine, Paris, France
[b] University of Warsaw, Warsaw, Poland

ARTICLE INFO

Article history:
Received 10 January 2017
Received in revised form 16 July 2018
Accepted 23 July 2018
Available online 25 July 2018

Keywords:
Proportional representation
Diversity
Multiwinner elections
Apportionment
Recommendation systems
Algorithms
Computational complexity
Approximation algorithms

ABSTRACT

We consider the following problem in which a given number of items has to be chosen from a predefined set. Each item is described by a vector of attributes and for each attribute there is a desired distribution that the selected set should have. We look for a set that fits as much as possible the desired distributions on all attributes. An example of application is the choice of members for a representative committee, where candidates are described by attributes such as gender, age and profession, and where we look for a committee that for each attribute offers a certain representation, i.e., a single committee that contains a certain number of young and old people, certain number of men and women, certain number of people with different professions, etc. Another example of application is the selection of a common set of items to be used by a group of users, where items are labelled by attribute values. With a single attribute the problem collapses to the apportionment problem for party-list proportional representation systems (in such a case the value of the single attribute would be a political affiliation of a candidate). We study the properties of the associated subset selection rules, as well as their computational complexity.

1. Introduction

Consider the following example. A research department has to choose k members for a recruiting committee. A selected committee should be gender-balanced, ideally containing 50% of male and 50% of female. Additionally, a committee should represent different research areas in certain proportions: ideally it should contain 55% of researchers specialising in area 1, 25% of experts in area 2, and 20% in area 3. Another requirement is that the committee should contain 30% of junior and 70% of senior researchers, and finally, the repartition between local and external members should be kept in proportions 30% to 70%. The pool of candidates from which the department can select members of such a committee is as shown in the table below.

In the given example, if the department wants to select $k = 3$ members, then it is easy to see that there exists no committee that would satisfy all the criteria perfectly. Nevertheless, some committees are better than others: intuitively we feel that in the selected committee the ratio of the numbers of members representing different genders should be either equal to 2:1 or to 1:2, the ratio of the numbers of members representing areas 1, 2 and 3, should be equal to 2:1:0. Further, the selected committee should contain one junior and two senior members, and exactly one member of the selected

☆ The preliminary version of this paper was presented at the *30th Conference on Artificial Intelligence (AAAI-2016)*.

E-mail address: p.skowron@mimuw.edu.pl (P. Skowron).

https://doi.org/10.1016/j.artint.2018.07.005

Name	Gender	Group	Age	Affiliation
Ann	*F*	1	*J*	*L*
Bob	*M*	1	*J*	*E*
Charlie	*M*	1	*S*	*L*
Donna	*F*	2	*S*	*E*
Ernest	*M*	1	*S*	*L*
George	*M*	1	*S*	*E*
Helena	*F*	2	*S*	*E*
John	*M*	2	*J*	*E*
Kevin	*M*	3	*J*	*E*
Laura	*F*	3	*J*	*L*

committee should have local affiliation. Such relaxed criteria can be achieved by selecting Ann, Donna, and George. Now, let us consider the above example for the case when $k = 4$. In such a case, the ideal ratios between the numbers of members for each of the four attributes should be equal to 1:1, 2:1:1, 1:3, and 1:3, respectively. Observe, however, that there exists no committee satisfying such relaxed criteria. According to different criteria, in this case the best committee can be for instance {Ann, Charlie, Donna, George}, with two externals instead of three, or {Charles, Donna, George, Kevin}, with males being over-represented.

In this paper we formalise the intuition given in the above example and we define what it means for a committee to be optimal, with respect to multi-attribute proportional representation. In our approach we leverage classical tools from political and social sciences, in particular we adapt the concept of *proportional apportionment* from the political science literature [3] to the case of multiple attributes. The central question of the apportionment problem is how to distribute parliament seats between political parties, given the numbers of votes cast for each party. Indeed, we can consider our multi-attribute problem, with the single attribute being a political affiliation of a candidate, and the desired distributions being the proportions of votes cast for different parties. In such a case we can see that selecting a committee in our multi-attribute proportional representation system boils down to selecting a parliament according to some apportionment criterion.

To emphasise the analogy between our model and the apportionment methods, we should provide some discussion on where the desired proportions for attributes come from. Typically, but not always, they come from *votes*. For instance, each voter might give her preferred value for each attribute, and the ideal proportions coincide with the observed frequencies. For instance, out of 20 voters, 10 would have voted for a male and 10 for a female, 13 for a young person and 7 for a senior one, etc.[1] It is worth mentioning that the voters might cast approval ballots, that is for each attribute they might define a set of approved values rather than pointing out the single most preferred one. On the other hand, sometimes, instead of votes, there are "global" preferences on the composition of the committee, expressed directly by the group, imposed by law, or by other constraints that should be respected as much as possible independently of voters' preferences.

There is a variety of apportionment methods considered in the literature (we refer the reader to the survey of Balinski and Young [3]). They are evaluated by means of properties; among those that are deemed important and have been extensively studied in the literature, we find *non-reversal*, *respect of quota*, *population monotonicity*, and *house monotonicity* (see [2]). We define the analogs of these properties for the multi-attribute domain. These properties give us some insights into the nature of multi-attribute committee selection mechanisms; in particular, their analysis allows us to view certain selection methods as generalisations of the appropriate apportionment rules. Specifically, following this approach, in this paper we define multi-attribute variants of the Hamilton rule and of the d'Hondt rule of apportionment, hereinafter referred to as the *multi-attribute Hamilton rule* and the *multi-attribute d'Hondt rule*.

The multi-attribute case, however, is also substantially different from the single-attribute one. In particular, multi-attribute proportional representation systems exhibit computational problems that do not appear in the single-attribute setting. Indeed, in the second part of our paper we show that finding an optimal committee is often NP-hard. However, we show that this challenge can be addressed by designing efficient approximation and fixed-parameter tractable algorithms. In particular, the core technical contribution of this paper lies in the analysis of approximation guarantees provided by the local-search algorithm for the problem of finding an optimal committee, with respect to a certain measure of multi-attribute proportional representation.

We believe that the model formalised in this paper has broad applications. As an example, consider a political system where the voters do not vote for the candidates directly, but rather for their opinions on various issues. For instance, quoting Lang and Xia [32], in 2012, voters in California had to decide in simultaneous multiple referenda whether to adopt each of the given eleven propositions[2]; a similar vote also took place in Florida. Given that the voters vote on propositions, our algorithms can be used to find a set of candidates that, in some sense, best represents opinions of voters about propositions. The number of propositions can be even larger: for instance, political parties have usually quite elaborate programs in which they refer to tens or hundreds of issues.

[1] How to aggregate in a consistent way ideal proportions specified by different voters is a nontrivial problem addressed in [15].

[2] http://en.wikipedia.org/wiki/California_elections,_November_2012.