ARTICLE IN PRESS

Discrete Applied Mathematics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Robust recoverable and two-stage selection problems

Adam Kasperski [a,*], Paweł Zieliński [b]

[a] Faculty of Computer Science and Management, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland

[b] Faculty of Fundamental Problems of Technology, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland

ARTICLE INFO

Article history:
Received 17 February 2016
Received in revised form 12 August 2017
Accepted 14 August 2017
Available online xxxx

Keywords:
Robust optimization
Computational complexity
Approximation algorithms
Selection problem

ABSTRACT

In this paper the following selection problem is discussed. A set of n items is given and we wish to choose a subset of exactly p items of the minimum total cost. This problem is a special case of 0–1 knapsack in which all the item weights are equal to 1. Its deterministic version has an $O(n)$-time algorithm, which consists in choosing p items of the smallest costs. In this paper it is assumed that the item costs are uncertain. Two robust models, namely two-stage and recoverable ones, under discrete and interval uncertainty representations, are discussed. Several positive and negative complexity results for both of them are provided.

1. Introduction

In this paper we wish to investigate the following SELECTION problem. Let $E = \{e_1, \ldots, e_n\}$ be a set of items. Each item $e \in E$ has a nonnegative cost c_e and we wish to choose a subset $X \subseteq E$ of exactly p items of the minimum total cost, $f(X) = \sum_{e \in X} c_e$, where $p \in [n] = \{1, \ldots, n\}$. This problem has a trivial $O(n)$-time algorithm which works as follows. We first determine in $O(n)$-time the pth smallest item cost, say $c_{(p)}$ (see, e.g., [11]), and then choose p items from E whose costs are not greater than $c_{(p)}$. SELECTION is a special, polynomially solvable version of the 0–1 knapsack problem, in which all the items have unit weights. It is also a special case of some other discrete optimization problems such as minimum assignment or a single machine scheduling problem with the weighted number of late jobs criterion (see [19] for more details). It can be seen as a basic resource allocation problem [18]. It is also a matroidal problem, as the set of feasible solutions is composed of all bases of an uniform matroid [29].

Suppose that the item costs are uncertain and we are given a *scenario set* $\mathcal{U}$, which contains all possible vectors of the item costs, called *scenarios*. We thus only know that one cost scenario $S \in \mathcal{U}$ will occur, but we do not know which one before a solution is computed. The cost of item $e \in E$ under scenario $S \in \mathcal{U}$ is denoted as c_e^S and we assume that $c_e^S \geq 0$. No additional information for scenario set $\mathcal{U}$, such as a probability distribution, is provided. Two methods of defining scenario sets are popular in the existing literature (see, e.g., [25]). In the *discrete uncertainty representation*, $\mathcal{U}^D = \{S_1, \ldots, S_K\}$ contains $K > 1$ explicitly listed scenarios. In the *interval uncertainty representation*, for each item $e \in E$ an interval $[\underline{c}_e, \overline{c}_e]$ of its possible costs is specified and $\mathcal{U}^I = \Pi_{e \in E} [\underline{c}_e, \overline{c}_e]$ is the Cartesian product of these intervals. The cost of solution X depends now on scenario $S \in \mathcal{U}, \mathcal{U} \in \{\mathcal{U}^D, \mathcal{U}^I\}$, and will be denoted as $f(X, S) = \sum_{e \in X} c_e^S$. In order to choose a solution two robust criteria, namely the *min–max* and *min–max regret* can be applied, which lead to the following two optimization problems:

$$\text{MIN-MAX SELECTION}: \min_{X \in \Phi} \max_{S \in \mathcal{U}} f(X, S),$$

* Corresponding author.
E-mail addresses: adam.kasperski@pwr.edu.pl (A. Kasperski), pawel.zielinski@pwr.edu.pl (P. Zieliński).

http://dx.doi.org/10.1016/j.dam.2017.08.014

$$\text{Min-Max Regret Selection}: \min_{X\in\Phi}\max_{S\in\mathcal{U}}(f(X,S)-f^*(S)),$$

where $\Phi = \{X \subseteq E : |X| = p\}$ is the set of all feasible solutions and $f^*(S)$ is the cost of an optimal solution under scenario S. The quantity $f(X,S) - f^*(S)$ is called a *regret* of solution X under scenario S. Both robust versions of the Selection problem have attracted a considerable attention in the recent literature. It turns out that their complexity depends on the way in which scenario set $\mathcal{U}$ is defined. It has been shown in [3] that, under the discrete uncertainty representation, Min–Max Regret Selection is NP-hard even for two scenarios. Repeating a similar argument as the one used in [3] gives the result that Min–Max Selection remains NP-hard even for two scenarios. Both problems become strongly NP-hard when the number of scenarios is a part of input [19]. Furthermore, in this case Min–Max Selection is also hard to approximate within any constant factor [19]. Several approximation algorithms for Min–Max Selection have been recently proposed in [13,19,20]. The best known, designed in [13], has an approximation ratio of $O(\log K/\log\log K)$. For the interval uncertainty representation both robust problems are polynomially solvable. The min–max version is trivially reduced to a deterministic counterpart, as it is enough to solve the deterministic problem for scenario $(\overline{c}_e)_{e\in E}$. On the other hand, Min–Max Regret Selection is more involved and some polynomial time algorithms for this problem have been constructed in [3,10]. The best known algorithm with running time $O(n \cdot \min\{p, n-p\})$ has been shown in [10].

Many real world problems arising in operations research and optimization have a two-stage nature. Namely, a complete or a partial solution is determined in the first stage and can be then modified or completed in the second stage. Typically, the costs in the first stage are known while the costs in the second stage are uncertain. This uncertainty is also modeled by providing a scenario set $\mathcal{U} \in \{\mathcal{U}^D, \mathcal{U}^I\}$, which contains all possible vectors of the second stage costs. If no additional information with $\mathcal{U}$ is provided, then the robust criteria can be applied to choose a solution. In this paper we investigate two well known concepts, namely *robust two-stage* and *robust recoverable* ones and apply them to the Selection problem. In the robust two-stage model, a partial solution is formed in the first stage and completed optimally when a true scenario reveals. In the robust recoverable model a complete solution must be formed in the first stage, but it can be modified to some extent after a true scenario occurs. A key difference between the models is that the robust two-stage model pays for the items selected only once, while the recoverable model pays for items chosen in both stages with the possibility of replacing a set of items from the first to the second stage, controlled by the recovery parameter.

Both models have been discussed in the existing literature for various problems. In particular, the robust two-stage versions of the covering [12], the matching [22] and the minimum spanning tree [21] problems have been investigated. The two-stage model has been also considered in the stochastic setting, i.e. when a probability distribution in scenario set is available. Namely, it has been applied to the minimum spanning tree [14], the 0–1 knapsack [23,24], the matching [22] and the maximum weighted forest [1] problems. The robust recoverable approach has been applied to linear programming [26], some network problems [6,7,28], the 0–1 knapsack [8], and recently to the traveling salesperson [9] and the minimum spanning tree [17] problems.

Our results. In Section 3 we will investigate the robust recoverable model. We will show that it is strongly NP-hard and not at all approximable, when the number of scenarios is a part of input. A major part of Section 3 is devoted to constructing a polynomial $O((p-k+1)n^2)$ algorithm for the interval uncertainty representation, where k is the recovery parameter. In Section 4 we will study the robust two-stage model. We will prove that it is NP-hard for two second-stage cost scenarios. Furthermore, when the number of scenarios is a part of input, the problem becomes strongly NP-hard and it has an approximability lower bound of $\Omega(\log n)$. For scenario set $\mathcal{U}^D$, we will construct a randomized algorithm which returns an $O(\log K + \log n)$-approximate solution with high probability. If $K = poly(n)$, then the randomized algorithm gives the best approximation up to a constant factor. We will also show that for the interval uncertainty representation the robust two-stage model is solvable in $O(n)$ time.

2. Problems formulation

Before we show the formulations of the problems we recall some notations and introduce new ones. Let us fix $p \in [n]$ and define

- $\Phi = \{X \subseteq E : |X| = p\}$,
- $\Phi_1 = \{X \subseteq E : |X| \leq p\}$,
- $\Phi_X = \{Y \subseteq E \setminus X : |Y| = p - |X|\}$,
- $\Phi_X^k = \{Y \subseteq E : |Y| = p, |Y \setminus X| \leq k\}, k \in [p] \cup \{0\}$,
- C_e is the deterministic, first-stage cost of item $e \in E$,
- c_e^S is the second-state cost of item $e \in E$ under scenario $S \in \mathcal{U}$, where $\mathcal{U} \in \{\mathcal{U}^D, \mathcal{U}^I\}$.
- $f(X,S) = \sum_{e\in X} c_e^S$, for any subset $X \subseteq E$.

We now define the two-stage model as follows. In the first stage we choose a subset $X \in \Phi_1$ of the items, i.e. such that $|X| \leq p$, and we add additional $p - |X|$ items to X, after observing which scenario in $\mathcal{U}$ has occurred. The cost of X under scenario S is defined as

$$f_1(X,S) = \sum_{e\in X} C_e + \min_{Y\in\Phi_X} f(Y,S).$$

Please cite this article in press as: A. Kasperski, P. Zieliński, Robust recoverable and two-stage selection problems, Discrete Applied Mathematics (2017), http://dx.doi.org/10.1016/j.dam.2017.08.014.