European Journal of Operational Research 256 (2017) 62–67

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Discrete Optimization

A new compact formulation for the discrete p-dispersion problem

David Sayah*, Stefan Irnich

Chair of Logistics Management, Johannes Gutenberg University Mainz, Jakob-Welder-Weg 9, Mainz D-55128, Germany

ARTICLE INFO

Article history:
Received 6 November 2015
Accepted 16 June 2016
Available online 18 June 2016

Keywords:
Facility location
Dispersion problems
Max–min objective
Integer programming

ABSTRACT

This paper addresses the discrete p-dispersion problem (PDP) which is about selecting p facilities from a given set of candidates in such a way that the minimum distance between selected facilities is maximized. We propose a new compact formulation for this problem. In addition, we discuss two simple enhancements of the new formulation: Simple bounds on the optimal distance can be exploited to reduce the size and to increase the tightness of the model at a relatively low cost of additional computation time. Moreover, the new formulation can be further strengthened by adding valid inequalities. We present a computational study carried out over a set of large-scale test instances in order to compare the new formulation against a standard mixed-integer programming model of the PDP, a line search, and a binary search. Our numerical results indicate that the new formulation in combination with the simple bounds is solved to optimality by an out-of-the-box mixed-integer programming solver in 34 out of 40 instances, while this is neither possible with the standard model nor with the search procedures. For instances in which the line and binary search fail to find a provably optimal solution, we achieve this by adding cuts to our enhanced formulation. With the new techniques we are able to exactly solve instances of one order of magnitude larger than previously solved in the literature.

1. Introduction

In the p-dispersion problem (PDP), we are given a set of candidate locations $I = \{1, 2, \ldots, n\}$ and an $n \times n$ matrix $(d_{ij})_{i,j \in I}$ with distances d_{ij} between facility i and j. The optimization task is to select $1 < p < n$ facilities from I such that the minimum distance between any pair of selected facilities is maximized.

In practice, this location problem occurs whenever a close proximity of facilities is less desirable. A standard application is concerned with the location of nuclear power plants. Therein, one is interested in minimizing the risk of losing multiple plants in the event that only one plant is subjected to an enemy attack. To achieve this, a selection of plants is desired so that interplant distances are as large as possible. Similar applications can be found in the military sector. In more peaceful contexts, one seeks for facilities of the same franchise system or for public facilities which have overlapping areas of service, e.g., schools, hospitals, waste collection plants, or electoral districts. We refer the reader to Kuby (1987) and to the comprehensive survey of Erkut and Neuman (1989) for an overview on the variety of applications of the PDP. Another area of application is recognized if distances are not interpreted physically but as a measure of the diversity between members of a group, e.g., products of the same portfolio (Saboonchi, Hansen, & Perron, 2014).

The contribution of this paper is a new compact formulation of the PDP. To highlight the main objective pursued with this model, note that we intend to provide a competitive exact approach for the PDP in which a major part of the overall optimization task is undertaken by an out-of-the-box software package. To make the new model competitive, it is delivered along with two simple enhancements: We exploit simple bounds on the optimal distance to reduce the size and to increase the tightness of the model. The bounds are obtained by very simple heuristics that are already known in the literature. We show that clique inequalities are valid for the new model and can be used to further strengthen it. For the separation of the clique cuts, we also suggest a greedy heuristic in order to keep the coding effort and the computational burden as low as possible.

We carry out computational experiments over large-scale test instances in order to compare the new formulation against a standard mixed-integer programming model. The enhanced formulation is solved to optimality by a mixed-integer programming solver in 34 out of 40 test instances, while this is not possible with the standard model. We also compare our enhanced model against two standard search procedures for the PDP, i.e., a line search and a binary search. This comparison is interesting because these search procedures are easy-to-implement and exact making use of the relationship between the PDP and the maximum independent set

* Corresponding author. Tel.: +4961313922087.
E-mail address: sayah@uni-mainz.de (D. Sayah).

http://dx.doi.org/10.1016/j.ejor.2016.06.036

problem. For instances in which either line or binary search, or both, fail to find a provably optimal solution, we achieve this by adding the clique cuts to our formulation.

The remainder of the paper is structured as follows: In Section 2, we present non-linear and linearized PDP formulations from the literature and introduce our new compact formulation that is based on exploiting the relationship between PDP and the independent set problem. Section 3 describes the setup of the computational study and its results. In the concluding Section 4, we briefly hint at potential algorithmic improvements and related dispersion problems for which the approach taken in this paper might be promising.

2. Formulations

Without loss of generality, we assume that the distance matrix (d_{ij}) is symmetric and that any non-diagonal value is strictly positive. All formulations are based on a graph representation of the problem. Let (I, E) be the complete graph in which locations I are the vertices and $E = \{(i, j) \in I \times I : i < j\}$ are the edges. Given any distance d, we further define subsets of edges as

$$E(d) = \{(i, j) \in E : d_{ij} < d\} \subseteq E.$$

The PDP is a bottleneck optimization problem with a max–min objective function (Hsu & Nemhauser, 1979). We now briefly review two existing non-linear formulations exploiting this fact before we present a standard mixed integer linear programming (MILP) model and our new formulation.

2.1. Non-linear formulations

The first formulation is the mixed integer non-linear program of Pisinger (2006): Define a vector of location variables $x = (x_i)_{i\in I}$ and let $x_i = 1$ indicate that candidate location $i \in I$ is open (0, otherwise). Using a continuous variable $d \geq 0$ for the minimum distance between open locations, the PDP can be written as

$$Z = \max d \tag{1a}$$

$$\text{s.t.} \sum_{i\in I} x_i = p \tag{1b}$$

$$d x_i x_j \leq d_{ij} \qquad (i, j) \in E \tag{1c}$$

$$x_i \in \{0, 1\} \qquad i \in I \tag{1d}$$

$$d \geq 0. \tag{1e}$$

The objective (1a) maximizes the minimum distance d, and exactly p candidate locations are opened because of (1b). The non-linear constraints (1c) impose that any two locations i and j are only opened simultaneously $(x_i x_j = 1)$ if their distance d_{ij} is at least d. The variable domains are given by (1d) and (1e).

The next PDP formulation utilizes the relationship between PDP and the maximum cardinality independent set problem in subgraphs $(I, \tilde{E})$ of the graph (I, E), which can be stated as follows:

$$\max \sum_{i\in I} x_i \tag{2a}$$

$$\text{s.t. } x_i + x_j \leq 1 \qquad (i, j) \in \tilde{E} \subseteq E \tag{2b}$$

$$x_i \in \{0, 1\} \qquad i \in I. \tag{2c}$$

A vector $x \in \{0, 1\}^I$ satisfying constraints (2b) is the incidence vector of a subset $S \subseteq I$ that contains pairwise non-adjacent nodes in the graph $(I, \tilde{E})$, i.e., $x_i = x_j = 1$ only if $(i, j) \notin \tilde{E}$. We refer to S as an *independent set (IDS)* of the graph $(I, \tilde{E})$.

For a given value of d, the set of feasible solutions to PDP with minimum distance at least d is given by

$$\mathcal{X}(d) = \left\{ x \in \{0, 1\}^I : \sum_{i\in I} x_i = p \quad \text{and} \quad x_i + x_j \leq 1 \ \forall (i, j) \in E(d) \right\}.$$

A vector $x \in \mathcal{X}(d)$ is the incidence vector of an IDS of size p in the graph $(I, E(d))$. This notation allows us to state the PDP in the form

$$Z = \max d \tag{3a}$$

$$\text{s.t. } \mathcal{X}(d) \neq \emptyset \tag{3b}$$

$$d \geq 0. \tag{3c}$$

The minimum distance d is maximized in (3a), while constraint (3b) states that a feasible choice of d has to ensure that an IDS of size p exists in $(I, E(d))$. We refer to the problem of deciding whether $\mathcal{X}(d)$ is non-empty for any d as the IDS problem. Erkut (1990) proposed another non-linear formulation similar to (3). Neither of the above non-linear formulations was supposed to be solved directly. The authors motivate the two subsequent categories of exact solution approaches to the PDP which can be found in the literature.

MILP-based approaches. These approaches are driven by compact linearized versions of model (1). We describe a standard MILP formulation of the PDP in Section 2.2. Some authors suggest to solve the compact model straightaway using any off-the-shelf MILP solver (Daskin, 1995; Kuby, 1987). Erkut (1990) tailored a branch-and-bound algorithm for the PDP.

Search procedures. Model (3) motivates a simple search algorithm, e.g., line or binary search, to find a largest minimum distance in combination with an efficient method to perform the feasibility tests in each iteration of the search. For a continuous version of the PDP defined on a tree, Chandrasekaran and Daughety (1981) propose a search procedure which requires consecutive solutions of anti-cover problems. The anti-cover problem (Chaudhry, McCormick, & Moon, 1986) and the d-separation problem (Erkut, 1990; Erkut, ReVelle, & Ülküsal, 1996) are synonyms for the maximum IDS problem. Pisinger (2006) suggests a binary search and considers cliques of size p for the feasibility test.

We position the contribution of this paper in the first category because a new compact formulation for the PDP is presented (see Section 2.3). Along with the new formulation, we provide in Section 2.3.2 a greedy but usually effective procedure to strengthen its linear relaxation by separating valid inequalities. In our computational tests, we benchmark the new formulation against the standard MILP formulation and against the search procedures known from the literature.

2.2. Kuby formulation

Using an appropriately large number M, a linearization of (1) suggested by Kuby (1987) can be written as

$$Z = \max d \tag{4a}$$

$$\text{s.t.} \sum_{i\in I} x_i = p \tag{4b}$$

$$d \leq M(2 - x_i - x_j) + d_{ij} \qquad (i, j) \in E \tag{4c}$$

$$x_i \in \{0, 1\} \qquad i \in I \tag{4d}$$

$$d \geq 0. \tag{4e}$$