European Journal of Operational Research 242 (2015) 21–33

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Discrete Optimization

Construction and improvement algorithms for dispersion problems

Roberto Aringhieri [a], Roberto Cordone [b,*], Andrea Grosso [a]

[a] Università degli Studi di Torino, Dipartimento di Informatica, Corso Svizzera, 185, 10149 Torino, Italy
[b] Università degli Studi di Milano, Dipartimento di Informatica, Via Comelico, 39, 20135 Milano, Italy

ARTICLE INFO

Article history:
Received 3 August 2013
Accepted 24 September 2014
Available online 7 October 2014

Keywords:
Combinatorial optimization
Dispersion problems
Binary quadratic programming
Tabu Search

ABSTRACT

Given a set N, a pairwise distance function d and an integer number m, the *Dispersion Problems* (*DPs*) require to extract from N a subset M of cardinality m, so as to optimize a suitable function of the distances between the elements in M. Different functions give rise to a whole family of combinatorial optimization problems. In particular, the *max-sum DP* and the *max-min DP* have received strong attention in the literature. Other problems (e.g., the *max-minsum DP* and the *min-diffsum DP*) have been recently proposed with the aim to model the optimization of equity requirements, as opposed to that of more classical efficiency requirements. Building on the main ideas which underly some state-of-the-art methods for the *max-sum DP* and the *max-min DP*, this work proposes some constructive procedures and a Tabu Search algorithm for the new problems. In particular, we investigate the extension to the new context of key features such as initialization, tenure management and diversification mechanisms. The computational experiments show that the algorithms applying these ideas perform effectively on the publicly available benchmarks, but also that there are some interesting differences with respect to the *DPs* more studied in the literature. As a result of this investigation, we also provide optimal results and bounds as a useful reference for further studies.

1. Introduction

Let N be a set of n elements, m a positive integer number smaller than n and $d : N \times N \rightarrow \mathbb{R}$ a distance function on the elements of N, such that $d_{ii} = 0$ for all $i \in N$, $d_{ij} \geq 0$ and $d_{ij} = d_{ji}$ for all $i, j \in N$. The literature denotes by *Dispersion Problems* (*DPs*) a family of problems which require to extract from N a subset M of cardinality m, so as to optimize a suitable function of the distances between the extracted elements (Erkut, 1990; Prokopyev, Kong, & Martinez-Torres, 2009). The natural mathematical programming formulations for these problems associate a binary variable x_i to each element $i \in N$ and set $x_i = 1$ if i belongs to M, $x_i = 0$ otherwise:

$$\max z = f_d(x) \tag{1a}$$

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

$$x_i \in \{0, 1\} \quad i \in N \tag{1c}$$

where notation $f_d\,(x)$ means that the objective is a composite function of vector x *through* the distance function d, and specifically it depends only on the distances d_{ij} between pairs of elements (i, j) such that $x_i = x_j = 1$.

A whole family of problems can be derived from (1) by specifying in different ways the expression of $f_d(\cdot)$. All of them share the same set of feasible solutions, but the properties and behaviour of their objective functions can be strongly different. In particular, the classical application of *DPs* has been the maximization of some dispersion index used as a measure of operational efficiency. This may refer to the location of facilities (Erkut & Neuman, 1989; Kuby, 1987), but also to the protection of biological diversity, the formulation of admission policies, the formation of committees, the composition of medical crews (Adil & Ghosh, 2005; Aringhieri, 2009; Glover, Kuo, & Dhir, 1998; Kuo, Glover, & Dhir, 1993; Weitz & Lakshminarayanan, 1998) and, more theoretically, the identification of densest subgraphs (Brimberg, Mladenovic, Urosevic, & Ngai, 2009). For example, the *max-sum DP*, more commonly known as *Maximum Diversity Problem* (*MDP*) aims to maximize the sum of the pairwise distances between all selected elements (Kuo et al., 1993):

$$f_d(x) = \frac{1}{2} \sum_{i,j \in N} d_{ij} x_i x_j \tag{2}$$

whereas the *max-min DP* aims to maximize the distance between the two closest elements (Erkut, 1990):

$$f_d(x) = \min_{i,j \in N : x_i = x_j = 1} d_{ij} x_i x_j \tag{3}$$

In contrast to the classical line of research, Erkut and Neuman (1991) and Prokopyev et al. (2009) introduced alternative definitions

* Corresponding author. Tel.: +39 0250316235; fax: +39 0250316373.
E-mail addresses: roberto.aringhieri@unito.it (R. Aringhieri), roberto.cordone@unimi.it (R. Cordone), andrea.grosso@unito.it (A. Grosso).

http://dx.doi.org/10.1016/j.ejor.2014.09.058
0377-2217/

of $f_d(\cdot)$ to express equity requirements, referring in particular to the idea of fairness among candidate sites for urban public facilities. This alternative approach is focused on an intermediate *aggregate dispersion* measure

$$D_i(x) = \sum_{j \in N} d_{ij} x_j \quad i \in N \tag{4}$$

that is the sum of the distances between each single element and the selected ones, and can be equivalently expressed as $\sum_{j \in M} d_{ij}$. Both papers investigate the optimization of the *max-minsum DP*, which aims to guarantee that each selected element is as distant as possible from the other ones, setting:

$$f_d(x) = \min_{i \in N: x_i = 1} D_i(x) \tag{5}$$

which is the minimum aggregate dispersion for the selected elements, and which should be maximized. Prokopyev et al. (2009) also consider the *min-diffsum DP*, which aims to guarantee that each selected element has approximately the same total distance from the other ones, setting

$$f_d(x) = \max_{i \in N: x_i = 1} D_i(x) - \min_{j \in N: x_j = 1} D_j(x) \tag{6}$$

that is the maximum difference between the aggregate dispersions of the selected elements. Notice that function $f_d(\cdot)$ should be minimized here, instead of maximized.

This work proposes heuristic algorithms for the *max-minsum DP* and the *min-diffsum DP*, inspired by the state-of-the-art methods for the *max-sum DP* and *max-min DP*.

Our first aim is to investigate whether the main ideas which guarantee a strong performance on efficiency-concerned *DP*s maintain their effectiveness when applied to equity-concerned *DP*s. On one hand, the identical structure of the feasible set for these two subfamilies of *DP*s suggests that it might be the case. On the other hand, the completely different structure of the objective function, and consequently of the so called *landscape* of the problem (Stadler, 1992), poses reasonable doubts on this assumption. See also Resende, Martì, Gallego, and Duarte (2010) for a discussion on the weak correlation between the optimal solutions of the *max-sum DP* and the *max-min DP*. From this perspective, we investigate the impact of some constructive heuristics based on the idea of determining a more favourable initial solution and compare them with the use of a random restart procedure as in Aringhieri and Cordone (2011).

The second aim of this work is to provide best known results for publicly available benchmark instances of equity-concerned *DP*s, thus stimulating further research on the topic, as done for the *max-sum DP* in Martì, Gallego, Duarte, and Pardo (2011). Since the instances considered in Prokopyev et al. (2009) are not publicly available and their size (50–100 elements) is currently too small for a significant algorithmic comparison, we adopted the benchmark instances of the *Optsicom* web site (http://www.optsicom.es/mdp). These were originally proposed for the *max-sum DP*, but can be directly employed for all *DP*s. Specifically, we consider instances up to 500 elements. For some of them, we also provide optimal results, or at least bounds, obtained with a general purpose Mixed Integer Linear Programming (*MILP*) solver.

The paper is organized as follows. Section 2 surveys the relevant literature. Section 3 presents the algorithms here proposed to solve the *max-minsum DP* and the *min-diffsum DP*, discussing in detail the basic ideas inspired by the literature on efficiency-concerned *DP*s. Section 4 reports and discusses the computational results. Appendix A (Tables A.1–A.6) reports the best known results on all the tested instances.

2. A survey on dispersion problems

Most of the literature on *DP*s concerns the *max-sum DP* and the *max-min DP*. Briefly summarizing, the exact methods for the *max-sum DP* can solve instances up to 100–150 elements (Aringhieri, Bruglieri, & Cordone, 2009; Erkut, 1990; Martì, Gallego, & Duarte, 2010; Pisinger, 2006), whereas larger instances require heuristic approaches. Most of these approaches are local search metaheuristics based on the simple exchange of elements in and out of the current solution. In particular, the hybrid metaheuristic proposed by Wu and Hao (2013) provides the best known results for a large set of benchmark instances, whose size goes up to 5000 elements. Other approaches exhibiting remarkable performances are Variable Neighbourhood Search (Aringhieri & Cordone, 2011; Brimberg et al., 2009), Iterated Tabu Search (Palubeckis, 2007), Learnable Tabu Search (Wang, Zhou, Cai, & Yin, 2012), basic Tabu Search (Aringhieri, Cordone, & Melzani, 2008; Duarte & Martì, 2007), Scatter Search (Aringhieri & Cordone, 2011; Gallego, Duarte, Laguna, & Martì, 2009) and *GRASP* (Duarte & Martì, 2007; Silva, de Andrade, Ochi, Martins, & Plastino, 2007). According to our experience on this problem, the key to impressively good performances and fast execution times is the use of strong, though not necessarily sophisticated, diversification mechanisms (Aringhieri & Cordone, 2011).

The *max-min DP*, on the other hand, suffers from a very flat landscape of the objective function (several different solutions have exactly the same value). This poses a severe challenge on local search metaheuristics, as discussed in Resende et al. (2010), where different heuristics are extensively compared, among which a *GRASP* with evolutionary path relinking exhibits the best performance. To partly overcome the issue of the flat landscape, this work minimizes also a secondary objective function, that is the number of pairs (i, j) in the solution such that d_{ij} is minimum. Della Croce , Grosso, and Locatelli (2009) reformulate the *max-min DP* as a dichotomic search on a sequence of instances of the *Maximum Clique Problem* (*MCP*), which are solved with the powerful Iterated Local Search (*ILS*) heuristic proposed in Grosso, Locatelli, and Pullan (2008). This approach allows to prove the optimality of the solution for several instances up to $n = 250$ elements, provided that the clique subproblem is solved with an exact algorithm. In the end, the Tabu Search algorithm described in Porumbel, Hao, and Glover (2011) applies separate add and drop operations to reduce the complexity of each iteration from quadratic to linear, and it adopts an extremely simple tabu rule: the drop operation always removes the oldest selected element. In this way, each element remains in the solution for exactly m iterations. The algorithm also exploits the sum of all pairwise distances between the elements of the solution as an auxiliary objective function to perturb the flat landscape of the problem.

Switching to equity-concerned models, the *max-minsum DP* has been introduced by Erkut and Neuman (1991) and the *min-diffsum DP* by Prokopyev et al. (2009), who provide *MILP* formulations and discuss the computational complexity of both problems, and of other related ones. They also apply a general-purpose solver on instances from 30 to 100 elements and a GRASP metaheuristic on the instances with 50 and 100 elements. This algorithm generates a starting solution adding one element at a time, chosen randomly from a restricted candidate list of random length, which includes the elements yielding the best partial solutions. Then, the starting solution is improved with a sort of first-improvement local search on a restricted neighbourhood. This attempts a random exchange between one element in the solution and one out of it: if the objective improves, the new solution is accepted and the search restarts from it. If it is rejected for a specified number of times, the improvement phase gives place to a new constructive phase. The whole algorithm terminates after a given number of constructive and improvement phases.

Some works on equity-concerned models (Prokopyev et al., 2009) allow the distance function d to assume also negative values. Notice that, due to the cardinality constraint, if all distances between two distinct elements are increased by a uniform amount $\delta > 0$, the value of each feasible solution correspondingly increases by a fixed amount depending on δ and on the cardinality m: