Knowledge-Based Systems 000 (2017) 1–13

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

An iterated local search algorithm for the minimum differential dispersion problem

Yangming Zhou[a], Jin-Kao Hao[a,b,*]

[a] LERIA, Université d'Angers, 2 Boulevard Lavoisier, 49045 Angers, France
[b] Institut Universitaire de France, 1 Rue Descartes, 75231 Paris, France

ARTICLE INFO

Article history:
Received 14 August 2016
Revised 1 March 2017
Accepted 30 March 2017
Available online xxx

Keywords:
Combinatorial optimization
Dispersion problems
Heuristics
Iterated local search
Three phase search

ABSTRACT

Given a set of n elements separated by a pairwise distance matrix, the minimum differential dispersion problem (*Min-Diff DP*) aims to identify a subset of m elements ($m < n$) such that the difference between the maximum sum and the minimum sum of the inter-element distances between any two chosen elements is minimized. We propose an effective iterated local search (denoted by ILS_MinDiff) for *Min-Diff DP*. To ensure an effective exploration and exploitation of the search space, ILS_MinDiff iterates through three sequential search phases: a fast descent-based neighborhood search phase to find a local optimum from a given starting solution, a local optima exploring phase to visit nearby high-quality solutions around a given local optimum, and a local optima escaping phase to move away from the current search region. Experimental results on six data sets of 190 benchmark instances demonstrate that ILS_MinDiff competes favorably with the state-of-the-art algorithms by finding 131 improved best results (new upper bounds).

1. Introduction

Let $N = \{e_1, e_2, \ldots, e_n\}$ be a set of n elements and d_{ij} be the distance between e_i and e_j according to a given distance metric such that $d_{ij} > 0$ if $i \neq j$ and $d_{ij} = 0$ otherwise. The minimum differential dispersion problem (*Min-Diff DP*) is to identify a subset $S \subset N$ of a given cardinality m ($m < n$), such that the difference between the maximum sum and the minimum sum of the inter-element distances between any two elements in S is minimized. Formally, *Min-Diff DP* can be described in the following way.

Let $\Delta(e_v)$ be the sum of pairwise distances between an element $e_v \in S$ and the remaining elements in S, that is:

$$\Delta(e_v) = \sum_{e_u \in S, u \neq v} d_{uv}. \tag{1}$$

The objective value f of the solution S is then defined by the following differential dispersion:

$$f(S) = \max_{e_u \in S}\{\Delta(e_u)\} - \min_{e_v \in S}\{\Delta(e_v)\}. \tag{2}$$

Then, *Min-Diff DP* is to find a subset $S^* \subset N$ of size m with the minimum differential dispersion, i.e.,

$$S^* = \arg\min_{S \in \Omega} f(S). \tag{3}$$

where Ω is the search space including all possible subsets of size m in N, i.e., $\Omega = \{S : S \subset N \text{ and } |S| = m\}$. The size of Ω is extremely large, up to a maximum number of $\binom{n}{m} = \frac{n!}{m!(n-m)!}$.

Min-Diff DP is one of many diversity or dispersion problems [1,30] which basically aim to find a subset S from a given set of elements, such that a distance-based objective function over the elements in S is maximized or minimized. These problems can be further classified according to two types of objective functions:

- *Efficiency-based* measures which consider some dispersion quantity for all elements in S. This category mainly includes the *maximum diversity problem* (MDP) and the *max-min diversity problem* (MMDP), which respectively maximizes the total sum of the inter-element distances of any two chosen elements and the minimum distance of any two chosen elements.
- *Equity-based* measures which guarantee equitable dispersion among the selected elements. This category includes three problems: (i) the *maximum mean dispersion problem* (Max-Mean DP) maximizes the average inter-element distance among the chosen elements; (ii) the *maximum min-sum dispersion problem* (Max-Min-sum DP) maximizes the minimum sum of the inter-element distances between any two chosen elements; (iii) the

* Corresponding author.
E-mail addresses: zhou.yangming@yahoo.com (Y. Zhou), jin-kao.hao@univ-angers.fr (J.-K. Hao).

http://dx.doi.org/10.1016/j.knosys.2017.03.028
0950-7051/

minimum differential dispersion problem considered in this work. It is worth noting that the cardinality of subset S is fixed except for Max-Mean DP.

In addition to their theoretical significance as NP-hard problems, diversity problems have a variety of real-world applications in facility location [21], pollution control [9], maximally diverse/similar group selection (e.g., biological diversity, admission policy formulation, committee formation, curriculum design, market planning) [17,27,28], densest subgraph identification [20], selection of homogeneous groups [8], web pages ranking [19,33], community mining [32], and network flow problems [7].

In this study, we focus on *Min-Diff DP*, which is known to be strongly NP-hard [30]. *Min-Diff DP* can be formulated as a 0–1 mixed integer programming problem (MIP) [30]. Thus it can be conveniently solved by MIP solvers like IBM ILOG CPLEX Optimizer (CPLEX). However, being an exact solver, CPLEX is only able to solve instances of small size (up to $n = 50$ and $m = 15$), while requiring high CPU times (more than 2500 s) [30]. To handle larger instances, heuristic and meta-heuristic algorithms are often preferred to find near-optimal solutions. In recent years, several heuristic approaches have been proposed in the literature [13,28,30]. In particular, in 2015, based on greedy randomized adaptive search procedure (GRASP), variable neighborhood search (VNS) and exterior path relinking (EPR), Duarte et al. proposed several effective hybrid heuristics [13]. Very recently (2016), Mladenović et al. proposed an improved VNS algorithm which uses the swap neighborhood both in its descent and shaking phases [28]. This new VNS algorithm significantly outperforms the previous best heuristics reported in [13] and is the current best performing algorithm available in the literature for *Min-Diff DP*. We will use it as our main reference for the computational studies.

Our literature review showed that, contrary to other diversity problems like MDP and Max-Min DP, for which many exact and heuristic methods have been investigated, there are only a very limited number of studies for *Min-Diff DP*. In this work, we partially fill the gap in terms of heuristic solving of *Min-Diff DP*, by proposing an effective heuristic algorithm for the problem (named ILS_MinDiff). We identify the main contributions of this work as follows.

First, in terms of algorithmic design, the proposed ILS_MinDiff algorithm is the first adaptation of the general three-phase search method to *Min-Diff DP* (Section 2). Compared to other meta-heuristic framework, this method has the main advantage of being conceptually simple and easy implementation. The proposed algorithm iterates through a descent-based local optimization phase, a local optima exploring phase, and a local optima escaping phase (Section 3). Like [13,28], the descent-based local optimization phase is based on the conventional swap neighborhood. Meanwhile, ILS_MinDiff distinguishes itself from the existing heuristic algorithms by at least two specific features. Its local optima exploring phase applies a weak perturbation technique (based on deterministic tournament selection) to discover nearby local optima around a given local optimum (Section 3.4). On the other hand, the local optima escaping phase calls for a strong perturbation technique (based on parametric random selection) to move away from deep local optimum traps (Section 3.5).

Second, in terms of performance, we present 131 improved best-known results (i.e., new lower bounds) out of 190 popular *Min-Diff DP* benchmarks ($\approx$ 69%). These improved lower bounds constitute a valuable contribution to *Min-Diff DP* studies since they can be used as new reference values to assess other *Min-Diff DP* heuristic algorithms. These tightened lower bounds could also be used within an exact algorithm for better bounding and possibly leading to an optimality proof of some benchmark instances. Together with the equal best-known results for other 42 instances, the computational outcomes confirm the relevance of the proposed algorithm.

Third, the availability of the source code of our ILS_MinDiff algorithm (see Section 4.2) contributes favorably to future research on *Min-Diff DP* and related problems. Specifically, the code can be used to perform comparative studies or solve other problems that can be formulated as Min-Diff DP. It can also serve as a key component (local optimizer) of more sophisticated *Min-Diff DP* algorithms.

The remainder of the paper is organized as follows. In the next section, we present a brief literature review on the iterated local search framework and its two recent variants. In Section 3, we describe the general framework and the key components of the proposed algorithm. In Section 4, we show an extensive experimental comparison between our algorithm and state-of-the-art algorithms. A parameter analysis is provided in Section 5, followed by conclusions in Section 6.

2. Related work on applications of iterated local search

As one of the most widely-used meta-heuristic approaches, *Iterated local search* (ILS) [24] has been successfully applied to solve a variety of combinatorial optimization problems. In spite of its conceptual simplicity, it has led to a number of state-of-the-art results. Fig. 1 shows that over the last two decades, there has been an increasing interest in ILS, as witnessed by the number of publications related to ILS.

ILS is a two-phase approach which explores iteratively the search zones around the last local optimum discovered by a local search procedure. Typically, an ILS algorithm repeats a local search phase to find a local optimal solution from a starting point and a perturbation phase to modify the local optimum to generate a new starting point for the next iteration (Algorithm 1). Based on the general ILS framework, several variants and extended approaches have recently been proposed, including two representative methods called breakout local search (BLS) [2,6] and three-phase search (TPS) [15]. The effectiveness of BLS and TPS have been verified on a variety of hard optimization problems and applications (see examples of Table 1). In the following, we present a brief review of these ILS variants.

Algorithm 1 Iterated local search.

```
S_0 ← GenerateInitialSolution()
S* ← LocalSearch(S_0)
while a stopping condition is not reached do
    S' ← Perturbation(S*, history)
    S*' ← LocalSearch(S')
    S* ← AcceptanceCriterion(S*, S*', history)
end while
```

Breakout local search introduced in [2,6] combines local search with a *dedicated and adaptive perturbation* mechanism. Its basic idea is to use a descent-based local search procedure to intensify the search in a given search region, and to perform dedicated perturbations to jump into a new promising search region once a local optimum is encountered. BLS is characterized by its adaptive perturbation. At the perturbation phase, BLS attempts to achieve the most suitable degree of diversification by dynamically determining the number of perturbation moves (i.e., the jump magnitude) and by adaptively selecting between several types of pre-defined perturbation operations of different intensities, which is achieved through the use of information from specific memory structures. As summarized in Table 1, BLS has reported excellent performances for several well-known combinatorial optimization problems. Algorithm 2 describes the general framework of BLS.

Please cite this article as: Y. Zhou, J.-K. Hao, An iterated local search algorithm for the minimum differential dispersion problem, Knowledge-Based Systems (2017), http://dx.doi.org/10.1016/j.knosys.2017.03.028