On solving max-mean dispersion to optimality

Michele Garraffa[1], Jérôme Malick[2], Frédéric Roupin[1]

[1] LIPN-CNRS UMR7030-Université Paris-Nord, France
{michele.garraffa,frederic.roupin}@lipn.univ-paris13.fr

[2] CNRS, Lab. J. Kunztmann, Grenoble, France
jerome.malick@univ-grenoble-alpes.fr

1 Introduction

This work deals with the so called *Max-Mean Dispersion Problem*, a combinatorial optimization problem whose aim is to find the subset of a set such that the average distance between the selected elements is maximized. Let us consider a set N formed by n elements. Given the dispersion values $d_{i,j} = d_{j,i}$ between elements $i, j \in N$, the problem consists in finding the subset $M \subseteq N$ with cardinality greater than or equal to 2, which maximizes the average value of $d_{i,j}$ for $i, j \in M$. The following fractional formulation, where each binary variable x_i indicates if the item $i \in N$ is selected or not, holds :

$$\max \quad \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} d_{i,j} x_i x_j}{\sum_{i=1}^{n} x_i} \tag{1a}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} x_i \geq 2 \qquad x_i \in \{0,1\} \ \forall i \in N \tag{1b}$$

Since the problem is strongly NP-hard and has many practical applications, several heuristics and exact approaches have been proposed. From a heuristic perspective, the best performing algorithm [4] is based on a fast evaluation of a simple one-flip neighborhood and has outstanding computational performances both considering solution quality and time. Moreover, the best exact method [1] is a branch-and-bound algorithm based on a semidefinite relaxation, which is able to solve to optimality instances up to 100 elements.

2 The proposed approach

In this work, we propose an exact approach that is an original mix of several operations research techniques. Our algorithm is in three phases detailed below, where the main numerical engine is the semidefinite-based binary quadratic optimization solver BiqCrunch [2]. The idea is to fully exploit the efficiency of the heuristic of [4] to reduce the problem to a classical binary quadratic optimization problem.

The algorithm consists of the following three phases :

1. generate a good candidate $\mathbf{x}^c$ for being the optimal solution ;

2. exclude a part of the solution space, in order to speed up the next phase ;
3. prove the optimality of $\mathbf{x}^c$.

More specifically, the first phase consists in executing a multistart variant of the heuristic in [4], in order to provide a good candidate to be the optimal solution. The second phase allows to find an integer value $\ell \geq 2$, such that $\sum_{i=1}^{n} x_i \geq \ell$ does not exclude the optimum. The value of ℓ is computed by solving easy k-cluster problems through the method proposed in [3]. Finally, the third phase proves the optimality of the solution $\mathbf{x}^c$.

Let $\epsilon > 0$ be a precision parameter and δ^c be the objective function of $\mathbf{x}^c$. The solution $\mathbf{x}^c$ is optimal if and only if the following binary quadratic program is not feasible :

$$\max \quad \sum_{i=1}^{n} x_i \tag{2a}$$

$$\text{s.t.} \quad \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} d_{i,j} x_i x_j \geq (\delta^c + \epsilon) \sum_{i=1}^{n} x_i \tag{2b}$$

$$\sum_{i=1}^{n} x_i \geq \ell \qquad x_i \in \{0,1\} \ \forall i \in N \tag{2c}$$

The aim is finding the maximum value for the cardinality of the subset, such that there exists a better solution than $\mathbf{x}^c$ (Constraint 2b).

3 Computational experiments

Some computational experiments have been performed on the instances of size $n = 100$ considered in [1]. They revealed two important facts :

- the proposed exact approach is around 100 times faster than the approach in [1] ;
- the heuristic in [4] computes all the optimal solutions in less than 1 second.

We are now working on extending the numerical experiments to larger instances with $n = 150$, that have not been solved to optimality yet.

Références

[1] M. Garraffa, F. Della Croce, F. Salassa, “An exact semidefinite programming approach for the max-mean dispersion problem", *Journal of Combinatorial Optimization*, 32, 1-23, 2016.

[2] N. Krislock, J. Malick, F. Roupin, “BiqCrunch : a semidefinite branch-and-bound method for solving binary quadratic problems", *To appear in ACM Transactions on Mathematical Software*, 2016.

[3] N. Krislock, J. Malick, F. Roupin, “Computational results of a semidefinite branch-and-bound algorithm for k-cluster", *Computers and Operations Research*, 66, 153-159, 2016.

[4] L. Xiangjing, H. Jin-Kao, “A tabu search based memetic algorithm for the max-mean dispersion problem", *Computers and Operations Research*, 72, 118-127, 2016.