Discrete Diversity Optimization

Rafael Martí, Anna Martínez-Gavara, Jesús Sánchez-Oro, and Sergio Pérez-Peló

Universitat de València
Dr. Moliner 50, 46100 Burjassot (Valencia), Spain
Rafael.Marti@uv.es, Gavara@uv.es

Universidad Rey Juan Carlos
Escuela Técnica Superior de Ingeniería Informática, Móstoles (Madrid), Spain
Jesus.SanchezOro@urjc.es, Sergio.Perez.Pelo@urjc.es

1 Introduction

The problem of maximizing diversity or dispersion deals with selecting a subset of elements from a given set in such a way that the distance among the selected elements is maximized. The definition of distance between elements is customized to specific applications, and the way that the overall diversity of the selected elements is computed results in different mathematical models.

Maximizing diversity by means of combinatorial optimization models has gained prominence in Operations Research (OR) over the last two decades, and constitutes nowadays an important area. In this talk, we review the milestones in the development of this area, starting in the late eighties when the first models were proposed, and identify three periods of time. The critical analysis from an OR perspective of the previous developments, permits us to establish the most appropriate models, their connection with practical problems in terms of dispersion and representativeness, and the open problems that are still a challenge. We also revise and extend the library of benchmark instances that has been widely used in heuristic comparisons. Finally, we perform an empirical comparison of the best and more recently proposed procedures, to clearly identify the state-of-the art methods for the main diversity models.

2 The three periods of research

We have identified three periods in the development of diversity and dispersion problems. The **early period**, from 1977 to 2000, where we can find the first models (MaxMin and MaxSum), and relatively simple algorithms to solve them, being the seminal paper by Kuby [5] in 1988 the origins of the area. Figure 1 shows a timeline with the first papers in the area.

In the second period, that we may call the **expansion period**, the first metaheuristics were proposed to target large instances effectively. Duarte & Martí [3] adapted both the Tabu Search and GRASP methodologies to the MaxSum model, triggering the interest of the metaheuristic community in this family of problems. Special mention deserves the work by Prokopyev et al. [8], where three new dispersion models were introduced: the MaxMinSum, the MaxMean, and the MinDiff. In this way, these authors clearly stated that there are different ways to model diversity maximization, opening many possibilities for future developments. This period lasts over a decade, ending with very efficient methods for some of the models, as shown in the empirical comparison of 30 methods by Martí et al. [7] performed in 2010, and with several solid research groups working on them. The boundaries defining the area of maximizing diversity were expanded with the inclusion of more realistic models built with capacity and cost constraints.

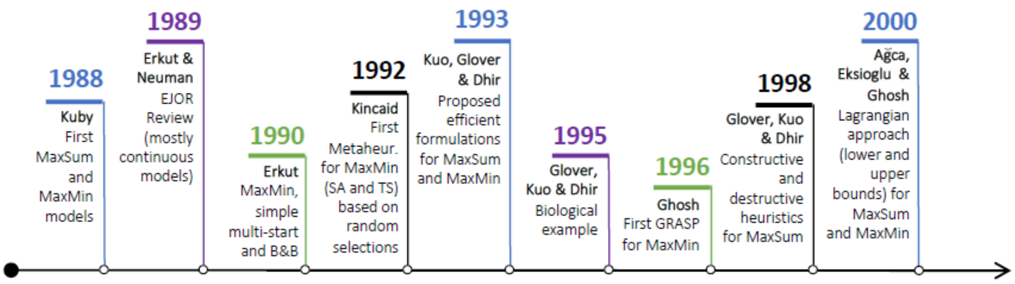

Figure 1: Early developments timeline.

The third period, that we call **the development period**, started in 2011 and is still in progress. From the heuristic side, the competition is now very high, due to the efficient methods published in the previous period, so only complex metaheuristics are proposed now. We devote now some attention to these methods for the most important models, the MaxSum and MaxMin, since they constitute the state-of-the art in discrete diversity.

2.1 The MaxSum model

Glover et al. [4] proposed four different heuristics for the MaxSum problem. The authors consider simple moves: constructive and destructive, that drive the search to approach and cross feasibility boundaries. These type of moves are natural in the maximum diversity problem, where the goal is to determine an optimal composition for a set of selected elements. The authors compare the solutions obtained with their heuristics with the optimal solutions in small instances, and conclude that the constructive method C2, and the destructive method D2 perform very well considering their simplicity.

The best method in the expansion period is the VNS by Brimberg et al. [1], originally devoted to the heaviest *k*-subgraph problem, which generalizes the MDP. The authors presented a basic VNS, called B-VNS that consists of three main elements. The first one, called Data Structure, allows the algorithm to efficiently update the value of the objective function; the second one, Shaking, generates solutions in the neighborhood of the current solution by performing random vertex swaps; and the third one is a local search procedure based on exchanges. Martí et al. [7] comparison shows that D2 and B-VNS are the best methods of their periods.

De Freitas et al. [2] proposed a Memetic Self-Adaptive Evolution Strategy, MSES. It is basically a population based algorithm that iterates over generations in which parents are mutated to produce children. A strength variable associated with each individual manages the mutation, and it is self-adjusted favoring that best configurations survive over time. As it is customary in memetic algorithms, the method includes a local search and a crossover, and as in previous implementations of the classic exchange-based local search, the authors propose 505 an efficient implementation based on splitting the move evaluation between the removed and the added contribution of its elements. The method is coupled with a tabu search that is selectively applied to the best children in the generation. The algorithm is implemented in Matlab, and it is compared with previous heuristics re-implemented in Matlab as well. The comparison on the MDPLIB instances favors the proposed method.

The last paper published so far on the MaxSum model is due to Zhou et al. [12], and it describes a memetic algorithm, called OBMA, improved with three search strategies:

- An opposition-based learning to reinforce population initialization.
- A tabu search to intensify the search in promising regions.
- A rank-based quality-and-distance pool updating maintains diversity in the population.

The opposition-based learning basically considers a candidate solution and its corresponding opposite solution. In the case of the MaxSum problem, the opposite solution is simply obtained by selecting some of the elements not selected in a given solution. The tabu search on the other hand, is based on a constrained swap strategy that manages the size of the explored neighborhood to speed up the method. As all the local search based methods for this problem, it is built upon a swap move that exchanges a selected with an unselected element in the solution. Finally, a rank-pool updating strategy decides whether an improved solution qualifies or not to enter into the population pool in which the memetic algorithm iterates. In particular, this strategy computes a score based on both quality and diversity to rank solutions in the updating process of the pool. In the next section we will compare these methods, not only to determine the best one, but also to evaluate the incremental contribution that each period of research is able to contribute to the quality of the solutions.

2.2 The MaxMin model

Resende et al. [11] applied the GRASP methodology to the MaxMin problem (MMDP), with an efficient implementation that is able to obtain high-quality solutions in short running times, outperforming all previous developments. An important characteristic of this GRASP for the MMDP is the definition of improving move. To efficiently search the flat landscape of the MaxMin, the authors introduced in the local search an extended meaning of the term improving, thus applying moves that result in solutions with the same objective function value but considered closer to the local optimal solution.

The GRASP method above is coupled with a Path Relinking (PR) post-processing for improved outcomes. The PR algorithm operates on a set of solutions, called elite set (ES), constructed with the best solutions obtained with GRASP. It basically creates paths of solutions between elite solutions. Let *x* and *y* be two solutions, PR starts with the first solution *x*, and gradually transforms it into the second one *y*, by swapping out elements selected in *x* with elements selected in *y*. The elements selected in both solutions *x* and *y* remain selected in the intermediate solutions generated in the path between them. The output of each PR iteration is the best solution, different from *x* and *y*, found in the path.

Porumbel et al. [9] proposed a fast local search for a model that combines the MaxMin and the MaxSum problems. In particular, the authors minimize the MaxMin objective function and consider the MaxSum as a secondary objective. The inclusion of this secondary objective is motivated by the fact that there may be a relative large number of solutions that qualify as optimal for the MaxMin, and it makes sense to choose the best one among them in terms of the MaxSum objective. This was already introduced in the very first paper published for these problems. Kuby (1988) introduced the MaxSum, the MaxMin, and what this author called a multi-criteria approach, arguing that the MaxSum model is an appropriate way to choose among the many alternate optima of the MaxMin problem.

3 Computational experiments

The benchmark instances for the diversity problem come from different sources that have been added over the years. The MDPLIB collects a total of 450 instances available at www.uv.es/rmarti/paper/mdp.html with a mirror server in www.optsicom.es/mdp. The library contains three sets of instances collected from different papers and named after their authors: GKD (Glover, Kuo, and Dhir), MDG (Martí, Duarte, and Gallego), and SOM (Silva, Ochi, and Martins). All the instances were randomly generated. We consider them in our experiments. We only report here two experiments, one for the MaxSum and one for the MaxMin model. We refer the reader to Martí et al. [8] for an exhaustive comparison and revision of the methods.

In the first experiment, we compare the best methods identified for each period time on the MaxSum model, namely D2, B-VNS, and OBMA, run with a time limit of 600 seconds per instance. We include in this experiment the solutions obtained with CPLEX and MSES which require on average about an hour of CPU time. Note that in many cases CPLEX is not able to certify the optimality, and we report its best feasible solution found (current lower bound when the time limit expires). This table shows the average percentage deviation from the best solution known (% dev), and the number of best solutions found (# best). Results are reported for each instance set. In the case of CPLEX, % dev is only reported in a set, when it obtains feasible solutions in all the instances in that set. The results in this table show that simple heuristics are not able to improve complex metaheuristics over a long period of time, and OBMA emerges as the best algorithm.

Table 1: Comparison of best heuristic methods for the MaxSum model.

	GKD-c	GKD-d	MDG-a	MDG-b	MDG-c	SOM-a	SOM-b	SOM-c	all
# inst.	20	140	60	60	20	50	20	80	450
% dev									
CPLEX	3.83	2.98	-	-	0.00	2.45	6.40	-	-
D2	10.05	24.99	20.03	18.44	76.23	26.92	19.26	19.48	29.92
B-VNS	0.00	0.00	0.01	0.01	0.02	0.00	0.00	0.05	0.01
MSES	0.00	0.71	1.15	0.72	0.41	0.00	0.07	0.89	0.49
OBMA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
# best									
CPLEX	0	79	3	5	0	18	0	0	105
D2	0	0	0	0	0	0	0	0	0
B-VNS	20	138	43	0	2	50	20	45	318
MSES	20	126	17	12	0	50	9	0	244
OBMA	20	138	60	60	20	50	20	80	447

0.00 means less than 0.001

Table 2: Comparison of best heuristic methods for the MaxMin model.

	GKD-c	GKD-d	MDG-a	MDG-b	MDG-c	SOM-a	SOM-b	SOM-c	all
# inst.	20	140	60	60	20	50	20	80	450
% dev									
CPLEX	4.55	0.22	-	-	-	0.00	15.00	-	-
C2Ad	56.06	91.35	65.38	98.16	100.00	68.61	35.00	100.00	76.82
D2Ad	16.01	42.41	44.51	74.54	75.23	62.53	35.00	86.09	54.54
GPR	4.00	30.46	7.87	54.02	100.00	7.30	10.00	64.21	34.73
DropAdd-TS	0.01	21.00	1.35	25.43	0.00	2.72	0.00	0.00	6.31
# best									
CPLEX	2	139	40	20	0	50	17	0	268
C2Ad	0	0	20	0	0	15	13	0	48
D2Ad	0	0	20	0	0	12	13	0	45
GPR	0	29	39	14	0	44	18	0	144
DropAdd-TS	20	32	58	39	20	47	20	80	316

0.00 means less than 0.001

The second experiment on the MaxMin model, reported in Table 2, shows that metaheuristics outperform heuristics, and DropAdd-TS arises as the best algorithm overall, with an average percentage deviation of 6.31% and 316 best solutions. CPLEX is able to obtain a total of 268 bests solutions out of 450 in the experiment, even improving the results achieve by GPR. This formulation solves to optimality many instances of large size (with n = 1000), and is able to obtain high quality lower bounds in even larger instances (n = 2000). Finally, comparing the two simple heuristics considered, we can see that the destructive method D2Ad obtains better solutions 1000 than the constructive one (C2Ad). Specifically, D2Ad presents an average deviation of 54:51% in contrast to the average deviation of 76:82% that C2Ad obtains.

4 Conclusions

In the early period (1980 - 2000) two mathematical models were proposed to capture the notion of diversity, the MaxSum and MaxMin, and simple heuristics were applied to solve these models in short computational times. During the last decade, called the development period (2010 - to now), researchers have been mainly working on the lines proposed in the previous decade. We want to give them credit because the competition among methods is now very hard, and the proposed methods both exact and heuristics are very sophisticated, but we believe that there is still some work to do on expanding the area.

The authors thank grant PID2021-125709OB-C21 by MCIN/AEI/ 10.13039/ 501100011033 and by ERDF A way of making Europe, and Generalitat Valenciana (CIAICO/2021/224).

References

[1] Brimberg, J., Mladenovic, N., Uro_sevi_c, D., & Ngai, E. (2009). Variable neighborhood search for the heaviest -subgraph. Computers & Operations Research, 36, 2885-2891.

[2] De Freitas, A., Guimaraes, F., Pedrosa Silva, R., & Souza, M. (2014). Memetic self-adaptive evolution strategies applied to the maximum diversity problem. Optimization Letters, 8, 705-714.

[3] Duarte, A., & Martí, R. (2007). Tabu search and GRASP for the maximum diversity problem. European Journal of Operational Research, 178, 71-84.

[4] Glover, F., Kuo, C.-C., & Dhir, K. S. (1998). Heuristic algorithms for the maximum diversity problem. Journal of Information and Optimization Sciences, 19, 109-132.

[5] Kuby, M. J. (1988). Programming models for facility dispersion: the p-dispersion and maxisum dispersion problems. Mathematical and Computer Modelling, 10, 792.

[6] Kuo, C.-C., Glover, F., & Dhir, K. S. (1993). Analyzing and Modeling the Maximum Diversity Problem by Zero-One Programming. Decision Sciences, 24, 1171{1185

[7] Martí, R., Gallego, M., Duarte, A., & Pardo, E. G. (2013). Heuristics and metaheuristics for the maximum diversity problem. Journal of Heuristics, 19, 591-615.

[8] Martí, R., Martínez-Gavara, A., Sánchez-Oro, J., and S. Pérez-Peló Discrete diversity and dispersion maximization. A review and an empirical analysis from an OR perspective. European Journal of Operational Research 299 (3) Invited review, 795 - 813 (2022).

[9] Porumbel, D. C., Hao, J. K., & Glover, F. (2011). A simple and e_ective algorithm for the MaxMin diversity problem. Annals of Operations Research, 186 , 27-{293.

[10] Prokopyev, O. A., Kong, N., & Martinez-Torres, D. L. (2009). The equitable dispersion problem. European Journal of Operational Research, 197, 59-67.

[11] Resende, M. G., Martí, R., Gallego, M., & Duarte, A. (2010). Grasp and path relinking for the max-min diversity problem. Computers & Operations Research, 37, 498-508.

[12] Zhou, Y., Hao, J. K., & Duval, B. (2017). Opposition-based memetic search for the maximum diversity problem. IEEE Transactions on Evolutionary Computation, 21, 731-745.