Expert Systems With Applications 108 (2018) 193–205

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Heuristics for the Bi-Objective Diversity Problem

J.M. Colmenar[a], R. Martí[b], A. Duarte[a,*]

[a] Department of Computer Sciences, Universidad Rey Juan Carlos, Móstoles, Madrid 28933, Spain
[b] Department of de Estadística e Investigación Operativa, Universidad de Valencia, Burjassot, Valencia 46100, Spain

ARTICLE INFO

Article history:
Received 13 October 2017
Revised 15 February 2018
Accepted 11 May 2018

Keywords:
Diversity problems
Multi-objective
Metaheuristics
Optimization

ABSTRACT

The Max-Sum diversity and the Max-Min diversity are two well-known optimization models to capture the notion of selecting a subset of diverse points from a given set. The resolution of their associated optimization problems provides solutions of different structures, in both cases with desirable characteristics. They have been extensively studied and we can find many metaheuristic methodologies, such as Greedy Randomized Adaptive Search Procedure, Tabu Search, Iterated Greedy, Variable Neighborhood Search, and Genetic algorithms applied to them to obtain high quality solutions. In this paper we solve the bi-objective problem in which both models are simultaneously optimized. No previous effort has been devoted to study the "combined problem" from a multi-objective perspective. In particular, we adapt the mono-objective methodologies applied to this problem to the resolution of the bi-objective problem, obtaining approximations to its efficient front. An empirical comparison discloses the best alternative to tackle this $\mathcal{NP}$-hard problem.

1. Introduction

The problem of maximizing diversity refers to the selection of a subset of elements from a given set in such a way that the diversity among the selected elements is maximized (Glover, Ching-Chung, & Dhir, 1995). We can find different applications of maximizing diversity in real-world situations. For instance, many universities in the United States, when determining admission policies, go beyond selecting through the academic grades, and also consider other factors in the search for a diverse set of students (Ramirez, 1979). In market planning, it is often desirable to maximize both, the number and the diversity of forces in a brand profile (Keely, 1989). Other contexts in which maximizing the diversity can be applied include plant cultivation (Porter, Rawal, Rachie, Wien, & Williams, 1975), social problems (Swierenga, 1977), immigration policies that promote ethnic diversity (McConnell, 1988), ecological preservation (Unkel, 1985), design of products (von Ghyezy, 1986), task-force management (Thomas, 1990), curriculum design (Jackson, 1991), and management of genetic resources (Glover, 1992).

As other optimization problems, in spite of its simplicity, it is a challenge even for modern solving techniques. As a matter of fact, the first obstacle that we encounter when dealing with diversity is its modeling. Note that when we talk about diversity, we are assuming the existence of a distance function in the space where elements belong. Distance functions, such as the well-known Euclidean distance, typically compute a value to measure the similarity or proximity of two elements. When we consider a subset of more than two elements, we have a distance value for each pair of elements in the subset, and we have to compute a single diversity value from all of them. In this way, we can compare two different subsets in our search for the best one. This is where the mathematical model plays a key role capturing the notion of diversity by specifying how to compute a single diversity value from many pairwise distances. A few models, and many solving methods have been proposed in the last few years.

The most studied model is probably the *Maximum Diversity Problem* (MDP) also known as the *Max-Sum Diversity Model* (Ghosh, 1996), in which the sum of the distances between the selected elements is maximized. Considering that n is the number of elements in the original set, and m the number of selected elements (the subset cardinality), we can formulate it in mathematical terms as:

$$\begin{aligned} \text{(MDP) Maximize} \quad & z_{MS}(x) = \sum_{i<j} d_{ij} x_i x_j \\ \text{s.t.} \quad & \sum_{i=1}^{n} x_i = m \\ & x_i \in \{0,1\} \qquad i = 1, \ldots, n. \end{aligned} \tag{1}$$

This formulation is based on the binary variables x_i indicating whether object i is selected or not. Note that although this math-

* Corresponding author.
E-mail addresses: josemanuel.colmenar@urjc.es (J.M. Colmenar), rafael.marti@uv.es (R. Martí), abraham.duarte@urjc.es (A. Duarte).

https://doi.org/10.1016/j.eswa.2018.05.013
0957-4174/

Table 1
Diversity measures according to Sandoya et al. (2018).

Measure	Mathematical function	Description
Sum	$\sum_{i<j, i, j \in M} d_{ij}$	This measure may address diversification among selected elements to distance.
Min	$\min_{i<j, i, j \in M} d_{ij}$	Focus on the minimum distance among the selected elements.
Mean	$\frac{\sum_{i<j, i, j \in M} d_{ij}}{\lvert M \rvert}$	Related to the Sum measure, is an average equity measure.
MinSum	$\min_{i \in M} \sum_{j \in M, j \neq i} d_{ij}$	This measure considers the minimum sum of distances, which corresponds to the aggregate dispersion among elements.
Diff	$\max_{i \in M} \sum_{j \in M, j \neq i} d_{ij} - \min_{i \in M} \sum_{j \in M, j \neq i} d_{ij}$	This measure can be understood as the difference between the largest and smallest values of the dispersion sum.

ematical model only contains one objective function and one constraint, it is indeed quite complicated to solve, due to the combination of a non-linear objective function with a discrete solution space (as a result of the binary variables). On the other hand, the combinatorial nature of its solutions makes its complete enumeration impracticable, thus being a challenge for both exact and heuristic solving methods.

The second model in terms of its popularity is probably the *Max-Min Diversity Problem* (MMDP) (Erkut, 1990), in which the minimum distance between the selected elements is maximized. It can be formulated in a similar way as follows:

$$\begin{aligned} &(\text{MMDP})\ \text{Maximize} && z_{MM}(x) = \min_{i<j} d_{ij} x_i x_j \\ &\text{s.t.} && \sum_{i=1}^{n} x_i = m \\ & && x_i \in \{0, 1\} \qquad i = 1, \ldots, n. \end{aligned} \tag{2}$$

The Max-Sum and Max-Min literature includes extensive surveys (Ağca, Eksioglu, & Ghosh, 2000; Erkut & Neuman, 1989; Kuo, Glover, & Dhir, 1993), exact methods (Ağca et al., 2000; Ghosh, 1996; Pisinger, 2006), and heuristics (Ghosh, 1996; Hassin, Rubinstein, & Tamir, 1997; Kincaid, 1992; Ravi, Rosenkrantz, & Tayi, 1994; Resende, Martí, Gallego, & Duarte, 2010). Although we can find other mathematical models to map the notion of diversity into functions and constraints, they did not receive much attention. Special mention deserves the work in Prokopyev, Kong, and Martinez-Torres (2009) in which four different models in the context of facility location and group selection are proposed. Table 1 summarizes the diversity measures known for a subset M of elements (Sandoya, Martínez-Gavara, Aceves, Duarte, & Martí, 2018). Note that all of them have been proposed in the context of single-objective optimization.

In this paper we study the diversity maximization from a multi-objective point of view. In particular, we consider the Max-Sum and the Max-Min measures as objectives to be simultaneously maximized. We have called this problem the Bi-Objective Diversity Problem (BODP). The aim of BODP is to provide the user with different solutions (subsets of the given set) with different objective values. To tackle the BODP we have explored six different methods from three families of algorithms: the Non-dominated Sorting Genetic Algorithm-II, usually known as NSGA-II (Deb, Pratap, Agarwal, & Meyarivan, 2002) and the second version of the Strength Pareto Evolutionary Algorithm, also known as SPEA2 (Zitzler, Laumanns, & Lothar, 2001), from the population-based algorithms; Greedy Randomized Adaptive Search Procedure (Feo & Resende, 1989) and Iterated Greedy (Ruiz & Stützle, 2007) from the construction-based algorithms; and Tabu Search (Glover & Laguna, 1997) and Variable Neighborhood Search (Hansen & Mladenovic, 2001) from the trajectory-based algorithms. We have adapted these six algorithms to the BODP proposing some new elements, such us the codification of solutions, new genetic operators, constructive methods and local search strategies. Given that those algorithms require several parameters to be adjusted, we have used iRace (López-Ibánez, Dubois-Lacoste, Cáceres, Stutzle, & Birattari, 2016) to automatically tune their configuration taking into account the hypervolume quality indicator. Finally, we have compared the six approaches in a well-known set of instances using three quality indicators for multi-objective optimization: hypervolume, set coverage, and epsilon indicator. The results show that the Tabu Search method obtains better results and spends shorter execution times.

The rest of the paper is organized as follows. In Section 2, we motivate the multi-objective nature of the problem. In Section 3, we describe the algorithmic proposals, while in Section 4, we detail the experimental experience. Finally, in Section 5, we draw the conclusions and the future work.

2. Problem motivation

Given that the main purpose of this paper is to maximize diversity, one may ask, why is it important? We can find numerous applications in the scientific literature to answer it. From a human resources perspective, promoting diversity and equity is a goal in many companies and organizations. This includes INFORMS, the American institute for operations research and management science. This institute even promotes an award, the WSC Diversity Award, to improve outreach and diversity among young researchers. It is nowadays a well-established principle in many companies that increasing human diversity in not only ethical but also beneficial for the company efficiency. This has been very-well summarized in the book by Scott Page (Page, 2007), where he states that “Diverse perspectives and tools enable collections of people to find more and better solutions and contribute to overall productivity”.

If we turn our attention to logistics and consider for example the area of vehicle routing, we can find specific models to create diverse routes. Martí, Velarde, and Duarte (2009) considers a multi-objective routing problem in which a set of paths from an origin to a destination must be generated. Finding different paths in a graph is a classical optimization problem, and in the context of hazardous materials transportation, we want to obtain spatially dissimilar paths that minimize the risk (distributing the risk over all regional zones to be crossed uniformly). The Path Dissimilarity Problem involves obtaining a set of p paths with minimum length and maximum diversity. This and related models have been extensively studied in Operations Research where maximizing diversity is an important point.

In this paper we focus on the two most studied models, Max-Sum and Max-Min, and propose a bi-objective model for an improved approach to identify diverse subsets in a given set. We target these two objectives for several reasons. The first one is that they are pure diversity models, while the others mentioned above are equity models according to Sandoya et al. (2018). These authors reviewed all the existing models and classified them according to their scope. A second reason to consider these two objectives is because of their relatively low correlation (Resende et al., 2010).