Information Sciences 230 (2013) 183–196

Contents lists available at SciVerse ScienceDirect

Information Sciences

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

An artificial bee colony algorithm for the maximally diverse grouping problem

Francisco J. Rodriguez [a,*], M. Lozano [a], C. García-Martínez [b], Jonathan D. González-Barrera [c]

[a] Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain
[b] Department of Computing and Numerical Analysis, University of Córdoba, Córdoba, Spain
[c] Department of Statistics, Operational Research, and Computer Science, University of La Laguna, San Cristobal de La Laguna, Spain

ARTICLE INFO

Article history:
Received 25 June 2012
Received in revised form 14 December 2012
Accepted 21 December 2012
Available online 8 January 2013

Keywords:
Maximally diverse grouping problem
Artificial bee colony algorithm
Iterated greedy

ABSTRACT

In this paper, an artificial bee colony algorithm is proposed to solve the maximally diverse grouping problem. This complex optimisation problem consists of forming maximally diverse groups with restricted sizes from a given set of elements. The artificial bee colony algorithm is a new swarm intelligence technique based on the intelligent foraging behaviour of honeybees. The behaviour of this algorithm is determined by two search strategies: an initialisation scheme employed to construct initial solutions and a method for generating neighbouring solutions. More specifically, the proposed approach employs a greedy constructive method to accomplish the initialisation task and also employs different neighbourhood operators inspired by the iterated greedy algorithm. In addition, it incorporates an improvement procedure to enhance the intensification capability. Through an analysis of the experimental results, the highly effective performance of the proposed algorithm is shown in comparison to the current state-of-the-art algorithms which address the problem.

1. Introduction

The maximally diverse grouping problem (MDGP) consists of partitioning a set of n elements (E) into m disjoint groups so that the diversity among the elements in each group is maximised. The diversity among the elements in a group is calculated as the sum of the dissimilarity values between each pair of elements. Let d_{ij} be the dissimilarity value between the elements i and j and $x_{ig} = 1$ if the element i is in the group g and 0 if not. The problem can thus be stated as:

$$\textbf{Maximise} \quad \sum_{g=1}^{m}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} d_{ij} \cdot x_{ig} \cdot x_{jg} \tag{1}$$

$$\textbf{Subject to} \quad \sum_{g=1}^{m} x_{ig} = 1 \quad i = 1,\ldots,n, \tag{2}$$

$$a_g \leqslant \sum_{i=1}^{n} x_{ig} \leqslant b_g \quad g = 1,\ldots,m, \tag{3}$$

$$x_{ig} \in \{0,1\} \; i = 1,\ldots,n; \quad g = 1,\ldots,m. \tag{4}$$

* Corresponding author.
E-mail address: fjrodriguez@decsai.ugr.es (F.J. Rodriguez).

http://dx.doi.org/10.1016/j.ins.2012.12.020

where constraint (2) ensures that each element is located in exactly one group and constraint (3) guarantees that the number of elements contained in group g is at least a_g and at most b_g. The MDGP also appears in the literature by other names, such as *the k-partition problem* in Feo et al. [12,13] and as *the equitable partition problem* in O'Brien and Mingers [36].

Different real-world applications of the MDGP can be found in the literature, covering a wide variety of fields. One of the most extended applications of the MDGP is to the assignment of students to groups [4,10,49–51,53]. The MDGP also arises in further fields such as the creation of diverse peer review groups [19], the design of VLSI circuits [13,49,50], the storage of large programs onto paged memory [32], and the creation of diverse groups in companies so that people from different backgrounds work together [6].

Several metaheuristic approaches were proposed to obtain high quality solutions to the MDGP. These include memetic algorithms [10,38], the tabu search [15], simulated annealing [38], and the variable neighbourhood search [38]. In this paper, we explore an alternative approach using the artificial bee colony (ABC) [5,24,26,27] method. The ABC algorithm is a new, population-based, metaheuristic approach inspired by the intelligent foraging behaviour of honeybee swarms. It consists of three essential components: food source positions, nectar-amount and three honeybee classes (employed bees, onlookers and scouts). Each food source position represents a feasible solution for the problem under consideration. The nectar-amount for a food source represents the quality of each solution (represented by an objective function value). Each bee-class symbolises one particular operation for generating new candidate food source positions. Specifically, employed bees search for food around the food source in their memory; meanwhile they pass their food information onto onlooker bees. Onlooker bees tend to select good food sources from those found by the employed bees, and then further search for food around the selected food source. Scout bees are a small group of employed bees that are transferred, abandoning their food sources and searching for new ones. Due to its simplicity and ease of implementation, the ABC algorithm has drawn much attention and has exhibited state-of-the-art performances for a considerable number of problems [28].

The rest of this paper is organised as follows. In Section 2 we provide an overview of particular metaheuristics that have previously been applied to the MDGP. In Section 3 we analyse the background to the ABC algorithm, which provides the basis of the specific algorithmic design we employ. In Section 4 we describe our ABC approach to the MDGP. In Section 5 we analyse the performance of the proposed ABC and draw comparisons with the existing literature. Finally, Section 6 contains a summary of results and conclusions.

2. Metaheuristics for the MDGP

In this section, we detail metaheuristics from the literature that are most relevant to the MDGP. The first approach to this problem was a *multistart algorithm* introduced by Arani and Lotfi [3]. This algorithm starts by generating a random solution that is then partially destroyed. Then, all the possible reconstructed solutions are explored by the algorithm in order to select the best one. This process is repeated until there is no improvement in the reconstructed solution.

Feo and Khellaf [13] presented several heuristics based on *graph theory*. Later, Weitz and Lakshminarayanan [50] carried out an experimental comparison of different heuristics for the MDGP and concluded that the best results were achieved by the *Lotfi–Cerveny–Weitz* [50] (LCW) method. This consists of an enhancement procedure that refines a random solution or a solution constructed by means of the *Weitz–Jelassi* [51] (WJ) algorithm. There are no significant differences in terms of the accuracy of the solutions between the two LCW variants; however, LCW with a random initial solution is considerably faster.

Fan et al. [10] proposed a *memetic algorithm* to deal with the MDGP (LSGA), combining a genetic algorithm with a local search procedure. The genetic algorithm uses a special encoding to group problems proposed by Falkenauer [8] and employs a procedure to initialise the population which ensures that the initial solutions conform to the constraints of the MDGP. LSGA employs a rank-based roulette-wheel strategy to select parents for the crossover operator. The latter is based on a special crossover operator proposed for grouping problems [8]. The local search procedure of LSGA implements a best improvement strategy based on exchanging elements between the groups [4]. Computational results show that LSGA is more effective than a pure genetic algorithm and LCW [50].

Later, Gallego et al. [15] developed a multistart metaheuristic for the MDGP that is based on the *tabu search methodology* (TS-SO). It consists of three main elements:

- A greedy constructive procedure to generate the initial solutions.
- A local search procedure based on the LCW method (T-LCW). This expands the LCW neighbourhoods to include insertions and adds a short-term tabu memory to prevent the recently moved elements from changing the group to which they belong for a number of iterations.
- A strategic oscillation method to explore solutions for which the group size limits may be amended. The strategic oscillation is coupled with the T-LCW improvement method, so that T-LCW is applied, relaxing lower and upper limits for all groups ($a_g = a_g - k$ and $b_g = b_g + k$). The value of k is reset to 1 after each successful application of T-LCW, otherwise k is increased by 1. In this way, the oscillation pattern is created. As this method works with unfeasible solutions, it is necessary to employ a repair mechanism. The repair mechanism consists of removing elements from groups that exceed their upper bounds and adding these elements to groups that fall below their lower bounds.