A Discrete Artificial Bee Colony Algorithm for the Multi-Objective Redistricting problem

Eric A. Rincón García, Antonin Ponsich, Roman A. Mora Gutiérrez, Pedro Lara Velázquez
Universidad Autónoma Metropolitana, Unidad Azcapotzalco, Dpto. de Sistemas
México D.F., C.P. 02200
aspo@correo.azc.uam.mx

Miguel A. Gutiérrez Andrade
Sergio G. De Los Cobos Silva
Universidad Autónoma Metropolitana, Unidad Iztapalapa, Dpto. de Ing. Eléctrica
México D.F., C.P. 09340
gamma@xanum.uam.mx

ABSTRACT

In this paper, the performance of two classical algorithms (simulated annealing and a discrete artificial bee colony) are compared on the redistricting problem, using a real example in Mexico and highlighting the superiority of the latter.

Categories and Subject Descriptors

G.1.6 [**Numerical analysis**]: Optimization; I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search—*Heuristic methods*

Keywords

Redistricting, Simulated Annealing, Artificial Bee Colony.

1. INTRODUCTION

The design of electoral zones arises when small geographical units (GU's) must be grouped into a predetermined number of districts [1], in such a way that one or several objective function(s) is (are) optimized while some restrictions, imposed by law, must be satisfied to guarantee democracy. This problem was proved to be NP-Hard [2]. Population equality, compactness and contiguity are typically regarded as essential in any democratic electoral process.

This study particularly focuses on a multiple objective electoral districting problem in Mexico and introduces a discrete artificial bee colony (DABC) algorithm to solve it. An adaptation of the initial algorithm, inspired by path relinking techniques, is proposed here for handling the discrete variables. The multi-objective feature of the problem is addressed through the common strategy based on a weighted linear aggregation function, such as that adopted by the Mexican Federal Electoral Institute (IFE). The resulting algorithm performance levels are compared with those of a tool used by the IFE, i.e. a classical Simulated Annealing (SA) algorithm.

2. DABC FOR REDISTRICTING PROBLEMS

Problem statement. As mentioned previously, population equality and compactness are important principles that should be promoted in the electoral district design. The corresponding objective functions are:

$$C_1(P) = \sum_{s \in S} \left(\frac{100 P_T}{d\left(\frac{P_N}{300}\right)} \right)^2 \left(\frac{P_s}{P_T} - \frac{1}{n} \right)^2 \tag{1}$$

$$C_2(P) = \sum_{s \in S} \left(\frac{PC_s}{4(AC_s)^{\frac{1}{2}}} - 1 \right) \tag{2}$$

Where P is a districting plan. P_N, P_T and P_s are population-related parameters (see [1] for a complete description). d is the maximum percentage of population deviation and $S = \{1, \ldots, n\}$ is the set of electoral districts that must be generated in the entity. PC_s and AC_s are the perimeter and the area of the considered district s, respectively. Finally, the fitness function to be minimized is defined according to an aggregation strategy: $\lambda_1 C_1(P) + \lambda_2 C_2(P)$, where $\lambda_1, \lambda_2 \in [0, 1]$ are weighting factors. In addition, the minimization is subjected to constraints that guarantee: (**R1**) each district is fully connected, (**R2**) the number of districts is equal to n and (**R3**) each GU is assigned to exactly one district.

The proposed discrete algorithm. Several proposals for extending the classical ABC working mode [3] to discrete search spaces exist in the specialized literature (see for instance [4]). In this work, an original modification based on the path relinking strategy is proposed to handle discrete decision variables.

The initial population of M food sources is generated randomly in such a way that each solution satisfies **R1-R3**. New solutions are created by employed bees using a path-relinking based approach: two solutions P_1 and P_2 are combined by randomly selecting a GU x. Thus, there is a zone Z_i in P_1 and a zone Z_j in P_2 such that $x \in Z_i$ and $x \in Z_j$. Let us consider the following sets: $K_1 = \{k : x_{ki} = 0, x_{kj} = 1\}$ and $K_2 = \{k : x_{ki} = 1, x_{kj} = 0\}$. Then, a percentage r ($r \sim \mathcal{U}(0,1)$) of GU's in K_1 are inserted into Z_i. The same percentage of GU's in K_2 are extracted from Z_i and inserted into any randomly chosen zone contiguous to Z_i. These moves can produce infeasible solutions, involving a repair process to enforce properties **R1-R3**.

Then, each onlooker bee selects a food source depending on its normalized fitness and a new food source is produced through the GU's insertion-extraction process explained above. The created food source replaces the former one through a

GECCO'12 Companion, July 7–11, 2012, Philadelphia, PA, USA.
ACM 978-1-4503-1178-6/12/07.

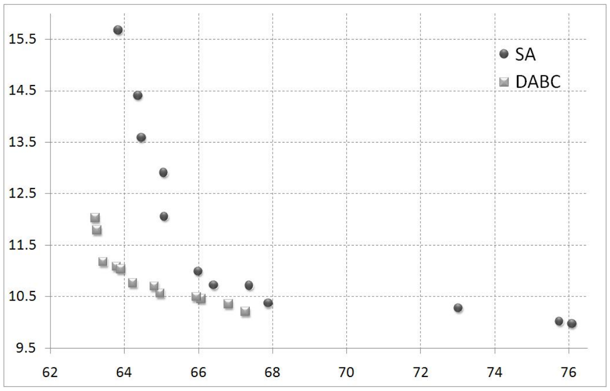

Figure 1: Pareto fronts for the two tested algorithms

greedy process. Finally, a similar process is implemented by the scouts to generate a new solution by combining a food source to be abandoned (after a defined number of cycles without improvement of this solution) and a solution chosen according to its normalized fitness. The new food source is accepted regardless its quality. The pseudocode of DABC is given in algorithm 1.

Algorithm 1: DABC Algorithm

```
Begin.
while Requirements are not met do
    Initialization.
    Produce new solutions through employed bees.
    Evaluate new solutions and select the best ones.
    Produce new solutions through onlooker bees.
    Evaluate new solutions and select the best ones.
    Abandonned solutions are replaced by scout bees.
    Memorize the best food source found so far.
end
```

3. COMPUTATIONAL EXPERIMENTS

Methodology. The SA and DABC algorithms were tested on the Mexican state of Mexico (13,096,686 inhabitants, 836 GU's and 40 districts to be created). In agreement with the Federal requirements, the value $d = 15\%$ (see equation 1) was used. The treatment of a single numerical instance may seem insufficient to provide robust conclusions but getting and processing government data represent harsh tasks, preventing us from solving more instances.

The set of weighting factors is defined as follows: $\lambda_1 = \{0.9, 0.8, 0.7, \ldots, 0.1, 0.01\}$, while $\lambda_2 = 1 - \lambda_1$. 10 independent executions of each algorithm were performed for each pair (λ_1, λ_2). The resulting solutions were subsequently filtered through a Pareto sorting procedure to identify the final non-dominated set. The algorithms working parameters were tuned in such a way that "convergence" (i.e., no-improvement) is achieved.

Results and discussion. The approximated Pareto fronts obtained with SA and DABC are illustrated in figure 1. Four performance metrics, typically used in multi-objective optimization (see [5] for more details), confirm the superiority of DABC. The global front, obtained by combining both techniques'solutions has 14 solutions, 12 from DABC and 2 from SA. The hypervolume metric is computed using as a reference the point with coordinates equal to the maximum values for each objective in any of the two obtained non-dominated sets. The results are 68.243 and 57.805 for DABC and SA, respectively. Regarding the set coverage metric, $\mathcal{C}(DABC, SA) = 83.33\%$ (83.33% of the efficient solutions produced by SA are dominated by at least one efficient DABC solution) while $\mathcal{C}(SA, DABC) = 0\%$. The Efficient Set Spacing metric is equal to 0.6518 and 0.1387 for SA and DABC, respectively, proving that the dispersion of the Pareto front achieved by DABC is thus much better than that of SA. Note, however, that DABC's non-dominated solutions are grouped together around the "knee" of the Pareto front but no solutions are found in the extreme regions, while the contrary is true for SA. Regarding the CPU time due to 100 executions, it was equal to 250 minutes for SA and 125 minutes for DABC.

Thus, in terms of convergence to an ideal front and distribution of the solutions over the Pareto front, DABC undoubtedly outperforms SA.

4. CONCLUSIONS

An original adaptation of the Artificial Bee Colony algorithm is proposed in the framework of combinatorial multi-objective optimization: the redistricting problem. This algorithm, as well as a classical Simulated Annealing procedure, were tested on a typical instance drawn from the Mexican electoral institute database. Their respective performance was evaluated in terms of convergence and dispersion of the resulting approximation of the Pareto front: the novel DABC algorithm produces better quality efficient solutions than SA, within lower running times. Two perspectives for future work might be treating other instances, in order to obtain robust conclusions on DABC performance levels, and exploring the reasons why DABC cannot identify efficient solutions in the extreme regions.

5. REFERENCES

[1] A., G.-A. M., and Rincón-García, E. A. Redistricting by square cells. *Lecture Notes in Computer Science 584* (2009), 669–679.

[2] Altman, M. Is automation the answer: The computational complexity of automated redistricting. *Rutgers Computer and Law Technology Journal 23* (1997), 81–141.

[3] Karaboga, D., and Basturk, B. A powerful and efficient algorithm for numerical function optimization: artificial bee colony (abc) algorithm. *Journal of Global Optimization 39* (2007), 459–471.

[4] Tasgetiren, M., Pan, Q., Suganthan, P., and Chen, A. A discrete artificial bee colony algorithm for the total flowtime minimization in permutation flow shops. *Information Sciences 181* (2011), 3459–3475.

[5] Zitzler, E., Deb, K., and Thiele, L. Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary Computation 8*, 2 (2000), 173–195.