Solving electoral zone design problems with NSGA-II. Application to redistricting in Mexico

Antonin Ponsich
Universidad Autónoma
Metropolitana, unidad Azcapotzalco
San Pablo #180
Mexico City, D.F. 02200
aspo@correo.azc.uam.mx

Eric A. Rincón García
Universidad Autónoma
Metropolitana, unidad Azcapotzalco
San Pablo #180
Mexico City, D.F. 02200
rigaeral@correo.azc.uam.mx

Román A. Mora Gutiérrez
Universidad Autónoma
Metropolitana, unidad Azcapotzalco
San Pablo #180
Mexico City, D.F. 02200
mgra@correo.azc.uam.mx

Sergio G. de-los-Cobos Silva
Universidad Autónoma
Metropolitana, unidad Iztapalapa
San Rafael Atlixco #186
Mexico City, D.F. 09340
cobos@xanum.uam.mx

Miguel A. Gutiérrez Andrade
Universidad Autónoma
Metropolitana, unidad Iztapalapa
San Rafael Atlixco #186
Mexico City, D.F. 09340
gamma@xanum.uam.mx

Pedro Lara Velázquez
Universidad Autónoma
Metropolitana, unidad Iztapalapa
San Rafael Atlixco #186
Mexico City, D.F. 09340
plara@xanum.uam.mx

ABSTRACT

The electoral zone design problem consists in redrawing the boundaries of legislative districts for electoral purposes, in such a way that federal or state requirements are fulfilled. In Mexico, both population equality and compactness of the designed districts are considered as two conflicting objective functions. The present work represents the first intent to apply a classical Multi-Objective Evolutionary Algorithm (the NSGA-II) to this hard combinatorial problem, whereas the Mexican Federal Electoral Institute has traditionnally used a Simulated Annealing (SA) algorithm based on a weighted aggregation function. Despite some convergence troubles, the NSGA-II obtains promising results when compared with the SA algorithm, producing better-distributed solutions over a wider-spread front.

KEYWORDS

Zone design problem, multi-objective optimization, NSGA-II

ACM Reference format:
Antonin Ponsich, Eric A. Rincón García, Román A. Mora Gutiérrez, Sergio G. de-los-Cobos Silva, Miguel A. Gutiérrez Andrade, and Pedro Lara Velázquez. 2017. Solving electoral zone design problems with NSGA-II.
Application to redistricting in Mexico. In *Proceedings of GECCO '17 Companion, Berlin, Germany, July 15-19, 2017,* 2 pages.
DOI: http://dx.doi.org/10.1145/3067695.3076103

1 PROBLEM DESCRIPTION

The zone design problem consists in aggregating small geographical units (GUs) into regions, in such a way that one (or more) objective function(s) is (are) optimized and some constraints are satisfied. Electoral redistricting is the best known case, due to its influence in the results of electoral processes and to its computational complexity, which has been proven to be NP-Hard [3]. In this framework, GUs are grouped into a predetermined number of zones or districts and democracy must be enforced through the satisfaction of some constraints imposed by law.

In Mexico, the National Electoral Institute (INE, acronym for *Instituto Nacional Electoral*) has used a Mathematical Programming model promoting the creation of districts accounting for the following three criteria.

Population equality. The state average population is determined dividing the number of inhabitants in the state by the number of districts to be formed. The number of inhabitants in each district must be close to the state average population and a maximum population deviation of 15% is allowed, formulated as an objective funtion to be minimized:

$$C_1(\mathcal{P}) = \sum_{s=1}^{n} \left(\frac{1 - \left(\frac{P_{Z_s}}{P_M} \right)}{0.15} \right)^2 \quad (1)$$

Where $\mathcal{P} = \{Z_1, Z_2, ..., Z_n\}$ is a redistricting plan with n districts, P_{Z_s} is the population of district Z_s ($s = 1, \ldots, n$) and P_M is the state average population.

Compactness. The redistricting process must promote the design of compact districts, that is, the boundaries of the districts must have a geometric shape as close as possible to the perimeter of a square having the same area. This concept is included as an objective function to be minimized:

$$C_2(\mathcal{P}) = \sum_{s=1}^{n} \left(\left(\frac{PC_{Z_s}}{\sqrt{AC_{Z_s}}} * 0.25 \right) - 1 \right) \quad (2)$$

Where PC_{Z_s} and AC_{Z_s} are the perimeter and the area of the considered district Z_s, respectively.

Contiguity. Districts must have geographic continuity taking into account the geo-electoral boundaries approved by the INE. This criterion is included as a hard constraint.

Different metaheuristics have been reported in the specialized literature for solving this problem, such as local search [4] or Evolutionary [1] and Swarm Intelligence [5] algorithms, while the INE

GECCO '17 Companion, Berlin, Germany

978-1-4503-4939-0/17/07…$15.00
DOI: http://dx.doi.org/10.1145/3067695.3076103

have used an improved Simulated Annealing (SA). However, to the best of our knowledge, no implementation of any Multi-Objective Evolutionary Algorithm (MOEA) has been tested for the treatment of the electoral zone design problem. Therefore, the primary purpose of this study is adapting and implementing an algorithm based on NSGA-II [2] for solving the electoral zone design problem.

2 COMPUTATIONAL EXPERIMENTS

The INE's implementation of SA is used here, involving a classical geometric decreasing cooling schedule. The NSGA-II implemented for this study is also the canonical version. Both techniques share the same initial solution generation and move/mutation mechanisms, adapted to the tackled problem. Crossover is performed through a Path Relinking wise strategy. The SA based algorithm uses a weighted aggregation function $f(\mathcal{P}) = \lambda_1 C_1(\mathcal{P}) + \lambda_2 C_2(\mathcal{P})$, with a set of 50 evenly distributed weight vectors λ. The solutions obtained by each run are filtered with a Pareto sorting procedure in order to get only the non-dominated front.

Computational experiments are carried out with the two algorithms for 8 out of the 32 Mexican states. 10 independent executions were performed for each algorithm. The best solutions of the 10 produced approximated Pareto fronts form the global front attained by each technique, while their combination is considered as an estimation of the true Pareto front $\mathcal{PF}_{true}$. The fronts are compared according to the well-known hypervolume and front coverage metrics, as well as the participation (proportion of solutions produced by a technique that participate to $\mathcal{PF}_{true}$).

The population size in NSGA-II is set to 50 individuals, in order to provide supposedly the same number of Pareto solutions as the SA. The NSGA-II was run during 7,000 generations because of time limitations. This means 0.35×10^6 objective function evaluations (OFEs) runs, taking about 3 hours each. In average, the number of OFEs used by SA is about 200 times higher than that assigned to NSGA-II, while NSGA-II needs about three times more time than SA to generate one single front. This behavior is due to the Pareto sorting procedure included in NSGA-II, which represents more than 97% of the total CPU time used by the NSGA-II.

First, solutions provided by each technique, the NSGA-II produces, as expected, 50 non-dominated solutions in each run while SA only finds, in average, slighlty more than 10 of them. Moreover, in almost all cases, one run of the MOEA determines a set of non-dominated solutions that covers a much broader space than that found by SA. NSGA-II is able to identify more solutions in the extreme parts of the front and its points are more evenly distributed than those of the SA.

Table 1: Participation $\mathcal{P}$ to $\mathcal{PF}_{true}$

State	% SA	% NSGA-II
Aguascalientes	37.93	**62.07**
Baja California	14.63	**85.36**
Baja Calif. Sur	37.50	**87.50**
Colima	36.84	**84.21**
Durango	**66.67**	33.33
Nayarit	5.56	**94.44**
Querétaro	26.67	**73.33**
Yucatán	7.41	**92.59**

Table 2: Front coverage metric (for global fronts).

State	C(SA, NSGA-II)	C(NSGA-II, SA)
Aguascalientes	**0.6604**	0.2143
Baja California	0.3518	**0.5714**
Baja Calif. Sur	**0.1111**	0.00
Colima	**0.2381**	0.00
Durango	**0.8889**	0.3333
Nayarit	0.00	**0.9091**
Querétaro	**0.7333**	0.3333
Yucatán	0.1613	**0.8889**

As a consequence, the composition of the (approximated) $\mathcal{PF}_{true}$ is biased in favor of NSGA-II, as indicated by Table 1. In five cases, more than 80% of the front is constituted by MOEA solutions.

Nevertheless, the MOEA experiences convergence troubles in some cases. Few SA non-dominated solutions, located near the knee region of the Pareto fronts, dominate a great number of the NSGA-II solutions. This observation therefore indicates that the NSGA-II, when trapped in a locally optimal front, cannot jump this barrier and getting to the real non-dominated front. Note, however, that this trend is not always reflected in the metrics since NSGA-II obtains better results than SA in 5 out of 8 for the front coverage, while results are mitigated for the hypervolume (Tables 3 and 2).

These preliminary computational experiments highlight the good behavior of the NSGA-II, which is able to identify a set of solutions well distributed over a wide-spread front. However, despite the diversity of solutions found, NSGA-II experiences convergence issues in the knee region of the front, promoting further investigation dealing with genetic operators, algorithm hybridization and regarding the computational efficiency of the Pareto sorting procedure.

Table 3: Hypervolumes compared between SA and NSGA-II

State	Global fronts		$\mathcal{PF}_{true}$
	NSGA-II	SA	
Aguascalientes	16.1176	**16.2443**	16.2499
Baja California	31.6089	**32.0449**	32.0534
Baja Calif. Sur	**0.0035**	0.0033	0.0035
Colima	**1.1276**	1.0919	1.1277
Durango	8.3585	**8.7821**	8.7838
Nayarit	**0.765**	0.7424	0.765
Querétaro	7.2341	**7.2381**	7.2471
Yucatán	18.9126	**19.0298**	19.042

REFERENCES

[1] C. Chung-I. 2011. A Knowledge-based Evolution Algorithm approach to political districting problem. *Computer Physics Communications* 182, 1 (2011), 209–212.

[2] K. Deb, A. Pratap, S. Aarwal, and T. Meyarivan. 2002. A fast and elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. *IEEE Transactions on Evolutionary Computation* 6, 2 (2002), 182–197.

[3] K. C. Gilbert, D. D. Holmes, and R. E. Rosenthal. 1985. A Multiobjective Discrete Optimization Model for Land Allocation. *Mgt Science* 31 (1985), 150–1522.

[4] F. Ricca and B. Simeone. 2008. Local search algorithms for political districting. *European Journal of Operational Research* 189, 3 (2008), 1409–1426.

[5] E. A. Rincón-García, M. A. Gutiérrez-Andrade, S. G. de-los Cobos-Silva, P. Lara Velázquez, R. A. Mora-Gutiérrez, and A. Ponsich. 2012. A Discrete Particle Swarm Optimization Algorithm for Designing Electoral Zones. In *Methods for decision making in an uncertain environment.* World Scientific Proceedings Series on Computer Engineering and Information Science (Eds.). Reus, Spain, 174–197.