International Journal of Advances in Computer Science & Its Applications
Volume 6 : Issue 1 *[ISSN 2250-3765]*
Publication Date : 18 April, 2016

A bi-objective proposal to group population without tap water or drainage network services

[María Beatríz Bernábe Loranca[1], Jorge Ruiz Vanoye[2], Rogelio González Velázquez[3], Martín Estrada Analco[4]]

***Abstract*—Proposing a multi-objective method is a common part of the process of supporting the decision-making process in population problems. This with the goal of motivating the selection of projects that help the population sector that has scarce water and drainage services at home. The method should provide a set of solutions that reflect the urgency to attend this population sector. Each solution is a set of geographical zone partitions formed by geographical units that fulfill the geometric compactness and homogeneity criteria, this homogeneity is given by the variable named "private houses without tap water, drainage network nor electricity services".**

We present basic theoretical aspects about the multi-objective methodology and a method supported on the order theory to find the Pareto Frontier. Given that the problem solves a kind of partitioning that optimizes two quality measures, the approximation to the optimal solution is made with variable neighborhood search (VNS). Once we obtained the set of partitions, understood as non-dominated solutions, we can analyze the results to show the relationship between the variables that must be attended in the population sectors with limited tap water and drainage network services. The study case that we deal with corresponds to the Metropolitan Zone of the Toluca Valley (ZMVT) considering its socio-economic data (Agebs) product of a census.

***Keywords*—grouping, multiobjective, population, pareto frontier.**

I. Introduction

The problems known as multi-objective problems are the ones that deal with the existence of multiple criteria to be fulfilled while in conflict with each other. This means that there are different solutions that could be chosen based on a series of opposite criteria. The decision making process (choosing a solution) resides in that according to a problem, the set of feasible points must be formed. Afterwards, a set of possible alternative solutions is tracked by any known technique. These possible solutions are the ones that satisfy the restrictions and preferences, which are executed over the proposed goals.

The method that we propose in this work generates a set of non-dominated solutions that follows basic aspects of the order theory; in particular we have taken ad-vantage of the partial order and non-comparable orders properties with the end of attaining a set of solutions that form the Pareto Frontier [1]. Said solutions are known as territorial groups (elements of the partition) characterized as geometrically compact and population-homogeneous for the "water" variables originated from a group of census variables.

Bernnábe[1], Gónzalez[3], Estrada[4]
Facultad de Ciencias de la Computación / Benemérita Universidad Autónoma de Puebla
México
beatriz.bernabe@gmail.com

Vanoye[2]
Universidad Autónoma del Estado de Hidalgo
México

On this point, the census data are useful to solve diverse structuring, planning, and organization problems for the population. However, one of the obliged tasks in this kind of problem is that decision making aspect, which obeys the current policies, available resources and established plans for social developments. Generally speaking, the problems that employ population data are directly related with the territorial design problems (TD) and their basic principle is that these must be analyzed in small groupings and with strategies in accordance to restrictions that describe the population. In this way, the solution proposals are focused on generating groups (territories) that satisfy restrictions demanded in TD such as geometric compactness and homogeneity for determined variables of interest (for example tap water and drainage network in our case). When a multi-objective partition is obtained, we have a number of well described and non-dominated groupings available in a Pareto Frontier. On the other hand, the information contained within the groups contributes to the duty of the decision maker to analyze the data regarding the viability, relation, location and analysis of the project. At this stage, our proposal creates compact-homogeneous groups according to population census indicators related to the lack of water and drainage services. The geographical objects that form the groups are known as Agebs and they are a product of the housing and population census of the INEGI (the national institute of statistics, geography and informatics of Mexico) and we have chosen the census data of the ZMVT to study them [2].

II. Problem statement

The goal is obtaining a set of groups of spatial data which composition is given by two components: geographical coordinates on the plane R^2 and a vector of census descriptive characteristics. The first component allows obtaining a distances matrix to process the geometric compactness calculus, one of the objective functions to minimize. The description vector is used to optimize the second objective function and consists in minimizing the heterogeneity of a given census variable. The selection of the variables will be made over the data that relates the population to the tap water and drainage network services.

According to the INEGI, their census data have allowed making diagnoses about sufficiency or deficit of services at home. These indicators have been the base to develop construction, expansion, improvement, financing and characterization strategies for houses. With the census variables it'll be possible to carry out studies about the deterioration of houses, the access to basic services that provide comfort, ease the domestic work and improve the quality of life.

A. Description

There's a physical search space for the geographic grouping. The geographic units are finite; this means each element is represented by its spatial location and an array of descriptive variables. The problem is discrete, combinatory, binary-integer and the aggregation of objects is made under the partitioning properties. To achieve compactness, we form the groups such that the geographic objects are geographically close to each other, by using an objective function that minimizes the sum of the distances between them. To achieve homogeneity, we seek equilibrium among the groups according to the total group value of the census variable under study. Having formed the groups under distance minimization, we calculate its homogeneity because in multi-objective problems the function to optimize has the same domain for all the objectives [3]. This is how we optimize compactness and homogeneity over the same partition. Then the best alternative is chosen regarding the m objectives. Mathematically speaking, there is a set *X* that is a subset of the space R^n such that $fi: X \rightarrow Rn$, $i=1,\ldots, m$, where *m* objectives exist.

III. Multi-objective

Definition 1. A multi-objective problem (MOP) can be defined for the minimization case as follows: Minimize f(x) given that $f: F \subseteq R^n \longrightarrow R^q, q \geq 2$ with feasible region in: $A = \{a \in F: gi(a) \leq 0, i = 1, \ldots, m\} \neq \emptyset$

The set A is called feasible region and we can say that the problem is subject to the restrictions gi: $R^n \longrightarrow R$ that can be any functions.

For R^n is possible to extend the concept by means of the following definition.

Definition 2. Given x, y vectors in R^n $x \leq y$ if and only if $x_k \leq y_k$ for every $k \in \{1, \ldots, n\}$ and $x < y$ if and only if $x \leq y$ with $x \neq y$, where $\leq$ is the usual order in R.

A. Pareto Frontier

A common option to use as a dominance relationship is known as the Pareto dominance defined as follows:

Definition 3. Given the multi-objective problem, minimize $f(x)$, where $f: F \subseteq R^n \longrightarrow R^q$ $q \geq 2$ with $A \subseteq F$ the feasible region. We say that a vector $x* \in A$ is non-dominated or a Pareto optimum, if there isn't a vector $x \in A$ such that $x < x*$.

Therefore, the answer to the problem of finding the best solutions (non-dominated solutions, however the dominance is defined in the technique) in a multi-objective problem is what is known as the solution set of the problem and the set of values of the objective function with a domain restricted to the vectors of the solution set (this is, the non-dominated vectors) is what we know as Pareto Frontier.

In this regard, thinking about the set of non-dominated vectors logically leads to the concept of partially ordered set.

Definition 4. The set $E(A; f)$ of Pareto efficient solutions (also known as set of Pareto optimums) is defined in the following way: $E(A, f) := \{a \in A: \nexists b \in A$ that fulfills $f(b) < f(a)\}$

This is the set of all the non-dominated vectors under the Pareto scheme.

A concept intimately related with the Pareto Frontier is the Pareto optimum concept. The Pareto optimum and Pareto frontier are the framework to work with in the decision making process for multi-criteria problems [1].

The set of Pareto optimums for a given multi-objective problem is a partially ordered set (poset) under a formal view. In the multi-objective problems we look for the minimal elements in the solution space Rn seen as a poset with the relationship $\leq$ given in definition 2.

As we are interested in the finding partitions of Ω (geographic units) that minimize the compactness and the heterogeneity we must make some minor adaptations to definitions 1, 3 and 4. For this we consider the collection of all the partitions $P = \{P: P \text{ is a partition of } \Omega\}$

Let $: P \longrightarrow R^2$. In our case the definition (1) is reduced to the following multi-objective problem: Minimize $f(P)$ given that $f: P \subset 2^{\Omega} \longrightarrow R^2$, with feasible region in $P = \{P \in 2\Omega: P \text{ is a partition of } \Omega\}$ where 2^{Ω} is the power set of Ω and $f(P) = (C(P), H(P))$ where C and H are the compactness and homogeneity functions f1 and f2 respectively, both with a domain in P and values in R.

We observe that the set of partitions P is generated from the finite set Ω then the image (Pareto Frontier) of the objective function f is finite and subsequently the Pareto Frontier is a discrete set.

According to the above, our problem can be expressed as follows:

First objective: Minimizing distances

This objective has been solved departing from a mono-objective partitioning algorithm [6]. Then, given a partition $P \in P$ for each $C \in P$ we randomly choose $c \in C$ and define the sum

$$S(P) = \sum_{C \in P} \sum_{i \in C} d(i, c)$$

Then the number $\min \{S(P): P \in P\}$ (1)

minimizes the intra-classes distances between geographic objects .

Second objective: Minimizing heterogeneity

Definition 5. Let $\Omega' = \{UG_1, UG_2, \ldots, UG_n\}$ be a set of n geographical units and $VC = \{X_1, X_2, \ldots, X_r\}$ a set of census variables that describe the UGs where each variable Xi is a function of the set of UGs in Ω' with values in the positive reals R^+. Given r intervals $I_j = [\alpha_j, \beta_j], j = 1, \ldots r$ and the characteristic functions $\chi[\alpha_j, \beta_j]: VC \rightarrow \{0,1\}$,

$$\chi[\alpha_j, \beta_j](X) = \begin{cases} 1 \text{ if } X \in [\alpha_j, \beta_j] \\ 0 \text{ in any other case} \end{cases}$$

Then we define the participation matrix associated to the group of UGs in Ω' with variables VC and conditions $I_j, j = 1, \ldots r$ as the matrix $M = (v_{ij})$ of size n×r where $v_{ij} = \chi[\alpha_j, \beta_j](X_j)\, X_j(UG_i)$. The matrix M contains all the

values of the variables that participate in the respective UGs. If $v_{ij} = 0$ we say that the variable X_j doesn't participate in the UG_i.

Having obtained the variables that participate in the grouping, we calculate the following, to homogenize the groups: for the variable under study we obtain an ideal average according to the number of groups, let's say that the variable of interest is X_j and that its ideal average is V_j, this happens when all the groups have the same value. However this isn't common in practice, then the real average for every group ($\frac{1}{n}\sum_{i=1}^{n} v_{ij}$)and we subtract the ideal average,

$$V_j - \frac{1}{n}\sum_{i=1}^{n} v_{ij} = \frac{1}{n}\sum_{i=1}^{n}(V_j - v_{ij}) \quad (2)$$

By minimizing this difference in the absolute value, we can obtain the cost of the objective function for homogeneity.

Minimize $y = f(x) = (f1(x), f2(x))$

f1: is the cost of minimizing the distances between UGs according to the equation (1) and f2 : is the cost of minimizing the heterogeneity of a census variable from the UGs according equation (2). Given the complexity of the problem, the optimization process consists of a VNS (descent) partitioning algorithm [4] whereas the minima is obtained by means of a non-comparable Pareto order relationship [5].

IV. VNS in multi-objective clustering

Variable neighborhood descent (VND) is obtained if all neighborhoods generated during the search process are explored completely [4].

Algorithm 1. Procedure VND (P, x, kvnd)

```
1 Select a set of neighborhood structures Nk: S →
P(S), 1 ≤ k ≤ kvnd;
2 Set stop = false;
3 repeat
4 Set k = 1;
5 repeat
6 x′ = BestImprovement(P, x,Nk(x));
7 if (f(x′) < f(x)) then
8 x = x′, k = 1; // Make a move.
9 else k = k + 1; // Next neighborhood.
10 endif
11 Update stop;
12 until (k == kvnd or stop);
13 until (f(x′) ≥ f(x) or stop);
14 return x.
```

The advantage of our mechanism to find better compromise solutions resides in the way that the grouping is solved: it returns a diverse set of partitions by using VNS. On the other hand, to find the subset of efficient and non-dominated solutions the mechanism evaluates each solution that is generated checking if it is non-dominated and non-comparable.

Each partition is represented by a pair of solutions composed of compactness and homogeneity (C, H). These solutions are checked under a variant of the Pareto dominance with the goal of obtaining a subset of non-dominated and non-comparable solutions (C, H). This subset is the Pareto Frontier.

The implicit partitioning must establish that every group is compact and that the sum of a certain population variable is as homogeneous as possible by means of a bi-objective function. The approximation to the optimum is made by combining VNS with a method supported on the order theory to find a set of non-dominated and non-comparable solutions through the minimal points that form the maxima set [5]. Finally, the algorithm process is described in an informal way as long as we consider that the initial solution is obtained considering a single objective (compactness) [6]

```
Algorithm 2. VND with Pareto Frontier filter
Procedure VND (P, x, kvnd)
1 Select a set of neighborhood structures Nk : S →
P(S), 1 ≤ k ≤ kvnd;
2 Set stop = false;
3 repeat
4 Set k = 1;
5 repeat
6 x′ = BestImprovement(x, LSit);
7 HandleMinimals(x′);
8 if (f(x′) < f(x)) then
9 x = x′, k = 1; // Make a move.
10 else k = k + 1; // Next neighborhood.
11 endif
12 Update stop;
13 until (k == kvnd or stop);
14 until (f(x′) ≥ f(x) or stop);
15 return x
```

The BestImprovement procedure takes the current solution x and the number of iterations LSit as input and chooses the best improvement found during the iterations made in this local search. Our neighborhood is defined by a pivot centroid which will remain in its position until the neighborhood is changed. A random exchangeable centroid will be chosen per iteration of the local search, and it'll be replaced by a randomly chosen non-centroid UG. After the exchange we reassign the UGs, each to their closest centroid to keep compactness. During this aggregation process we add the distances between the UGs and their centroids. For the homogeneity we add the value of the census variable for each UG and the total of both is returned. Besides the number of iterations we have included another stopping criterion for the local search: If the new solution generated in any iteration dominates the input solution, according to our Pareto order relationship, then the local search ends returning this non-dominated solution.

The procedure "HandleMinimals" is described next.

```
Algoritmo 4. HandleMinimals(x)
1 if(sizeOf(minimals)>0)
2  for(i = 1) to (i == sizeOf(minimals))
3    if(isDominated(minimals[i], x))
4      remove(minimals, i)
5      add(minimals, x)
6      break for
7    else
8      break for
9    end if
10 end for
11 end if
12 add(minimals, x)
13 return minimals
```

The algorithm HandleMinimals implements our Pareto order relationship to determine if the new solution generated "x'" dominates any of the current members of the minimal solutions list, if it does, the solution "x'" will replace the solution it dominates in the minimals list. If the list is empty,

"x"' is simply added to it. The final step returns the minimals list.

On this point, the non-dominated elements possess properties such as the maxima set (set of minimals or maximals) [5]. The properties of this set have been important to identify the non-dominated solutions in our method.

V. VNS in multi-objective clustering

We focus in offering a set of groups of population related to tap water and drainage network services. These groups are formed by spatial objects known as Agebs. Each Ageb consists of a vector of variables, which represent geo-statistical data from a census.

An important factor is making a decision about the viability of assigning resources to vulnerable population sector. We have chosen the population with partial tap water and drainage network services as study case. This means that the variables described on the list below are the ones that participate in the grouping and to make sure all of the Agebs are involved; we have filtered the variables with values between 0 and 100%.

Assuming that there's a government program to support this population sector, they need a set of groupings that show the distribution of this population regarding the following census variables, which are strongly related with each other:

Z119 Total of inhabited houses.

Z120 Inhabited private houses.

Z137 Private houses with drainage connected to the septic tank, ravine, crack, river, lake or sea.

Z138 Private houses without drainage network service.

Z140 Private houses with tap water inside the building.

Z141 Private houses with tap water within the land.

Z142 Private houses with tap water available by transportation (public water tap or from another home)

Z143 Private houses that only have tap water and drainage network services.

Z144 Private houses that only have drainage network and electricity services.

Z145 Private houses that only have tap water and electricity services.

Z146 Private houses that have tap water, drainage network and electricity services.

Z147 Private houses that don't have tap water, drainage network nor electricity services.

Z136 Private houses with drainage connected to the public network. [2]

A. *Application description*

The model we have described has been applied to a socio-economic problem, where we assume that 8, 32, 64 and 100 compact groups are required where the Agebs are very close to each other to ease the transportation implicit in social welfare programs that will attend the population with neglected water related services. The groups must be formed by Agebs with values on their drainage network and tap water services variables. The homogeneity we want is over private houses that don't have tap water, drainage nor electricity services (Z147).

The VNS parameters are set to 15 local search iterations and 2 runs over the whole set of neighborhood structures.

The homogeneity value was stable for all the tests and the execution time has an average of 60 seconds.

In each graph the x axis represents the compactness versus the y axis which is the homogeneity. If the decision maker is interested in sacrificing compactness or homogeneity he must choose a solution from the marked points. The values in each table associated to the graph are the minimum costs for both objectives. With the goal of testing that the solutions obtained with our method correspond to non-dominated solutions, we have employed Nodom, a software that filters the non-dominated data from an input vector [7].

We show in figure 1 and table 1 an example for 32 groups (g). The first column in table 1 belongs to the cost of the Homogeneity function (H) and the second one, to the compactness cost (C). These solutions are the set of non-dominated solutions that our algorithm generates, which we have improved. In a previous work we could see some dominated solutions close to the Pareto Frontier [8].

Finally, we have gathered the Pareto Frontiers (PF) from different tests into one graph. We used over 14 variables related with tap water and drainage network services, trying to balance the homogeneity in the variable Z147 (Private houses that don’t have tap water, drainage network nor electricity).

TABLE I. 11 SOLUTIONS

64194	634.9375
1464701	451
817700	514.5625
897517	504.5625
816472	562.5625
773141	629.75
797208	570.125
980917	492.5625
1013870	489.375
787619	580.125
793255	570.5625

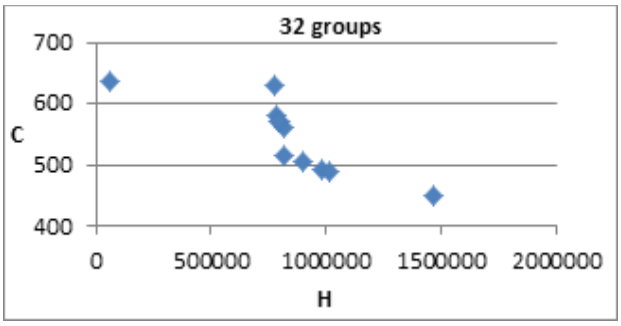

Figure 1. FP for 32 g

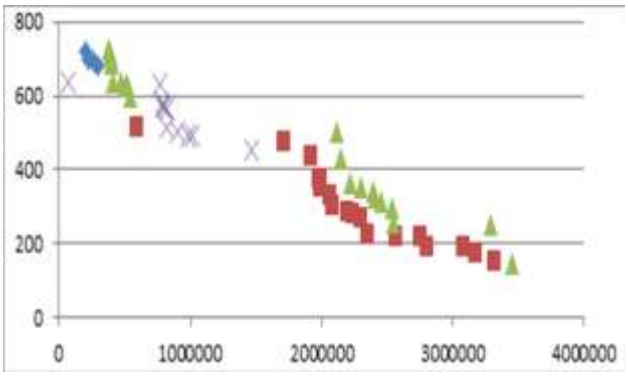

Figure 2. Blues → 8g, greens →32g, reds →64g, stars → 100g

VI. Conclusions

Our contribution is a partitioning around the medoids algorithm in a multi-objective context with VNS that surpasses a previous algorithm even though it reutilizes core procedures. However, a diversification in the neighborhood structures has been implemented resulting in a Pareto Frontier with a better distribution of solutions along the frontier line. Furthermore, our algorithm has interesting properties because it produces the Pareto Frontier by means of a reliable method, analogous to the properties of the minimals in a Hasse diagram, where these minimals are non-dominated and non-comparable points. The efficiency of the minimal solutions that our algorithm generates is tested when the unfiltered solutions are fed to Nodom, a software that filters non-dominated solutions.

On the other hand, the majority of grouping algorithms with a single optimization measure can work well but only for a certain amount of data, or some lack robustness concerning the variations in the cluster form and uniformity besides the proximity to the optimum.

In this work, we have proposed an alternative approach: simultaneously optimizing two objectives with VNS for a clustering problem for spatial data, however, our algorithm can group other kinds of data.

We have shown that our approach offers robustness in the selected solutions that form the Pareto frontier but we still need to deal with a problem that concerns many multi-objective researchers: the real Pareto frontier.

References

[1] Lara L; (2003) "Un estudio de las Estrategias Evolutivas para problemas Multiobjetivo, Tesis de Maestría en Ciencias en la especialidad de Ingeniería Eléctrica Opción Computación, Codirectores Dr. Carlos A. Coello Coello y Dr. Alin Carsteanu, Cinvestav IPN.

[2] INEGI. (2000). Sistema para la consulta de información censal 2000, (SCINCE), XII Censo General de Población y Vivienda 2000. Retrieved from http://www.inegi.org.mx/prod_serv/contenidos/espanol/catalogo/Default.asp?accion=2&upc=702825496371.

[3] Ríos D. (2008) "Sobre soluciones optimas en problemas de optimización multiobjetivo", Trabajos de Investigación Operativa. Editor Springer Berlin/ Heidelberg, ISSN 0213-8204, 2 (1) 49-67.

[4] Hansen P.; Mladenovic, N. (2003) "Variable neighbourhood search", In Fred Glover and Gary A. Kochenberger editors, Handbook of Metaheuristics, Kluwer

[5] Kung, H.; Luccio, F.; Preparata, F. (1975) "On Finding the Maxima of a Set of Vectors", Journal of the ACM (JACM), 1975, v.22 n.4, pp, 469-476.

[6] Bernábe Loranca Beatriz, Espinosa Rosales José E., Ramírez Rodríguez Javier, Osorio Lama, María A., "A Statistical comparative analysis of Simulated Annealing and Variable Neighborhood Search for the Geographic Clustering Problem", Computación y Sistemas, vol. 14, núm. 3, 2011, pp. 295-308, Instituto Politécnico Nacional, Distrito Federal, México.

[7] Nodom (2007) http://www.cs.cinvestav.mx/~emoobook/nodom/nonodom.html

[8] María Beatríz Bernábe Loranca, David Pinto Avendaño, Jorge A. Ruiz-Vanoye, José Espinosa Rosales, Elías Olivares Benitez. "A multi-objective proposal for the aggregation of economically inactive population; International Journal of Combinatorial Optimization Problems and Informatics, Vol. 3, No. 1, Jan-April 2012, pp. 70-79. ISSN: 2007-1558)