Lecture Notes in Management Science (2012) Vol. **4**: 202–211
4th International Conference on Applied Operational Research, Proceedings

ISSN 2008-0050 (Print), ISSN 1927-0097 (Online)

A hybrid metaheuristic for the partitioning problem with homogeneity constraints on the number of objects

María Beatriz Bernábe Loranca[1,2], David Pinto Avendaño[1], Elias Olivares Benitez[2], Rogelio González Velázquez[1], J. L. M. Flores[2] and J. R. Vanoye[2]

[1]*Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla, Puebla, México*
dpinto@cs.buap

[2]*Posgrado de Logística y Dirección de la cadena de Suministro, Universidad Popular Autónoma del Estado de Puebla, Puebla, México*
elias.olivares@upaep.mx

Abstract. Partitioning is a combinatorial optimization problem that has been widely discussed because of two main aspects: its computational complexity and its application to diverse kind of problems. In addition, partitioning has been used on the solution of some clustering issues that arise when dealing with territorial design problems. Besides the particular properties of the classic partitioning, in some cases, it is needed to consider additional constraints that need to be solved, such as the homogeneity of the number of objects that made up the final clusters. In this paper, we present a solution applied to partitioning problems having the moderate balance constraint, by determining thresholds which allows it obtaining a well defined range on the dimensionality of the groups. The aforementioned problem with constraints is considered to have a high computational complexity and, therefore, in order to obtain approximate solutions for this problem, we have incorporated a particular hybrid metaheuristic of partitioning that combines simulated annealing and variable neighborhood search. We present the problem, a solution proposal, the mathematical model and a set of experimental tests for samples taken from a study case based on 469 geographical data of the Metropolitan Zone of the Toluca Valley.

Keywords: compactness; hybrid metaheuristic; homogeneity; partitioning

Introduction

Territorial design is the problem of bringing together *n* small geographical areas into *m* bigger groups called territories, so that those territories satisfy the particular requirements of the problem in question. Many are the territorial design problems involved in enterprises or governmental institutions demands which have geographical data such as spatial units or geographical areas. Some examples of such problems are the alignment of selling territories, electoral districts, among others. Geometric compactness, continuity, connectedness and homogeneity are some of the most demanded constraints. Territorial design is a combinatorial problem which is NP hard, therefore, requiring the use of heuristic methods for obtaining approximate solutions Tavares et al (2007). One of the reasons of such computational complexity is the requirement of grouping geographical units under spatial constraints Zoltners and Shina (1983).

In this paper, it is proposed a solution to the compactness and moderate homogeneous grouping for the number of objects that made up the groups. The clustering is carried out by means of a partitioning algorithm based on medoids that minimizes distances between objects and centroids, whereas the homogeneity for the number of objects is calculated according to the variable bounds demanded by the required homogeneity. Due to the computational complexity of the problem tackled in this paper, a hybrid metaheuristic of simulated annealing (SA) with Variable Neighborhood Search (VNS) has been used. These heuristics have proved their efficiency on difficult problems of combinatorial optimization.

As study case, we use 469 basic geographical statistical areas (AGEBs) from the Metropolitan Zone of the Toluca Valley. Instances from 2 to 50 groups are solved for experimental purposes.

The rest of this document is organized as follows: Section 2 discusses the partitioning problem. Section 3 explains the hybrid metaheuristic designed for the problem solved in this paper. The computational tests are shown in Section 4. Finally, the conclusions are given.

Partitioning

Cluster analysis is a methodology that can be adapted to different problems requiring data grouping. Hierarchical clustering and Partitioning are two different categories of cluster analysis. There exist evidence showing that partitioning algorithms have the ability of creating groups even when exist spatial conditions on the dataset. Some encouraging results using partitioning around medoids and dynamic clouds with geometrical compactness constraints have been reported Bernábe et al (2011). However, a big challenge consists on extending the algorithm with homogeneity constrains under the number of objects that made up the groups.

In order to include the homogeneity constrain into partitioning algorithms, it is necessary to set this aspect as a primary restriction in a combinatorial optimization model. It is convenient to preserve the ideas originally obtained in the solution of the compact partitioning and basic concepts of the well known *k*-means method and dynamic clouds, such as the fact that each class may be represented by some object regularly named the kernel or nucleus (average point, an individual, a group of individuals or the class or a set of parameters). The underlying idea is that given a set of nucleus, the remaining individuals are assigned to the closest nucleus conforming in this way the classes, so that it may proceed with the calculus of the new nucleus with the classes conformed in the previous step. The final step is an iteration of the previous steps until stabilization is obtained. The method starts with an initial configuration of nucleus, and it may be proven that it converges to some partition in which the criterion given cannot be improved. The criterion is defined based on the context and the nucleus type Trejos et al (1990). For the purpose of this paper, we have developed a partitioning algorithm based on the idea of dynamic clouds and medoids as nucleus Kaufman and Rousseeuw (1987).

The model proposed

This model has been implemented using the basic properties of partitioning algorithms and additionally, because of the high computational complexity of the problem, we have incorporated a hybrid metaheuristic of Simulated Annealing Kirkpatrick et al (1983) and the technique of Variable Neighborhood Search Mladenović and Hansen (1997).

Homogeneity-compactness model

The model of combinatorial optimization of Territory Design Problem is:

Let $\mathrm{T}=\{\mathrm{BGU}_1,\mathrm{BGU}_2,\ldots\mathrm{BGU}_\mathrm{n}\}\subseteq \mathrm{R}^2$ a territory of basic geographical statistical areas BGAs. $C=\{c_1,c_2,\ldots,c_n\}$ set of centroids where $c_i \in BGU_i$ and $c_i=(x_i,y_i)\ \forall i=1,2,\ldots n$ with x-longitude and y-latitude. Let $d_{ij}=d(c_i,c_j)$ Euclidian distance from crentoide *i* to *j*. let G_i a subset of C called group. Agebs are indivisible units and belong to groups. Groups are represented by centroids. The problem of partitionin of *T* is to find a partitioning $P=\{G_1,G_2,\ldots,G_p\}$ so that

$$\min z=\sum_{i=1}^{p}\sum_{j\in G_i} d_{ij}.$$

The constraints are:

$$G_i \neq \varnothing,\ \forall i=1,2,..,p;$$

$$G_i \cap G_j=\varnothing,\ \forall i\neq j;$$

$$\bigcup_{i=1}^{p} G_i=T.$$

The minimization of z ensure the compactness, but in this case, it is neccesary that the groups must be as homogeneous as possible considering the cardinality of each group. So we introduced two parameters of tolerance: lower bound (LB) α_p and upper bound for each p (UP) β_p for each p, i.e.,

$$\alpha_p \le |G_i| \le \beta_p,$$

where

$$\alpha_p = (1-k)\ |T|/p \quad , \beta_p = (1+k)\ |T|/p$$

and k is the percentage of tolerance.

The hybrid metaheuristic solution proposed

In mathematical programming, the term heuristic is applied to a procedure of resolution for optimization problems. The heuristic is a procedure for which exists a high degree of confidence of finding high quality solutions with reasonable computational costs, even if the optimality of feasibility is not guaranteed. There exist the cases in which the closeness to the global optimum may not be established.

Some heuristics for solving optimization problems may be more generic or specific than others; however, the specific heuristic methods must be designed particularly for the given problem, using all the available information and the theoretical analysis of the model. These procedures use to have a higher performance than the generic heuristics which, on the other hand, present some advantages such as their robustness, adaptability and simplicity, besides that they may improve their performance by using other computational resources.

The metaheuristic term may be understood as the one with a superior level of the heuristic because it is a smart strategy of high performance for improving generic heuristic procedures in an optimization problem.

The common metaheuristics refer to relaxation methods, constructive process, neighborhood searches and evolutionary procedures; but the searching heuristics are the ones that constitute the central paradigm of these techniques of resolution for optimization problems. The hybrid metaheuristic proposed in this paper is based on the searching technique, incorporating a variable neighborhood search on the convergence of simulated annealing.

Searching metaheuristics guide the procedures by means of transformations or movements in order to go through the space of alternative solutions and, iteratively exploit the neighborhood structures associated to the initial solutions.

Monotonic search, hill climbing or local search algorithms contain particular rules for solving problems; in particular, the last one obtains improvements based on the analysis of similar solutions, using the concept of neighborhood solutions. Local search analyzes the neighborhood solutions around the current solution. The problem with this search strategy is that it usually stops finding local optima instead of a global one, there comes the importance of including metaheuristics for extending the local search process to be a global searching.

This aspect has suggested the construction of the hybrid algorithm that we present in this paper. Simulated annealing helps local search to escape from local optima. When the final number of iterations are reached, the convergence is obtained, and the suboptimal solution obtained with simulated annealing is given as initial solution to the neighborhood search variable procedure.

Simulated annealing (SA)

This method of neighborhood searching is characterized by a criterion of acceptance of neighborhood solutions self-adapted through its execution. It uses Temperature (T), a value that determines if a worse solution than the current may be accepted. The temperature starts at a high value which is named initial Temperature (To), and it is reduced in each iteration by means of a cooling parameter, $\alpha(\cdot)$, until the final Temperature (Tf), is reached. In each iteration, a number of neighborhood solutions, L(T), are generated. The exact number of solutions generated in each iteration may be fixed a priori or set it variable. For each neighborhood solution, the acceptance criterion must be applied in order to verify if the current solution is replaced by this neighbor. If the neighbor is better than the current solution, then it is replaced automatically, however, if the neighbor is worse than the current solution, still the chance that it may be replaced with certain level of probability, which allows the algorithm to escape from local optima. The level of probability for accepting a worse solution depends on the difference of costs between the current solution and its neighbors, δ, and also of the temperature T: $P_{acceptance} = \exp(-\delta/T)$.

The higher is the temperature, the better is the probability of accepting worse solutions. In this way, the algorithm accepts solutions much more worse at the beginning of the execution (exploration step), but at the end of the execution it rarely accepts worse solutions (exploitation step). On the other hand, the smaller is the cost difference, the better is the probability of accepting worse solutions. After generated the L(T) neighborhood solutions (at the end of each iteration), the temperature is decreased in order to proceed with the next iteration.

The algorithm in pseudocode is given as follows:

```
INPUT (To, α, L(t), Tf)
T← To Initial value for the control parameter
Sact←Generateinitial solution
WHILE T >= Tf DO Stopping condition
BEGIN
FOR cont← 1 TO L(t) DO Cooling speed (T)
BEGIN
Scand← Select solutionN(Sact) Creation of a new solution
☆←cost(Scand)-cost(Sact) Computation of cost difference
IF U(0,1) < e^(-☆/T) OR ☆<0 Application of acceptance criterion-
THEN Sact←Scand
END
T←α(T) Cooling mechanism
END
{Write as solution the best of the visited Sact}
```

Basic variable neighborhood search (VNS)

Basic VNS is a strategy that alternates local search with random movements on the neighborhood structures which vary on systematic manner.

The steps for basic VNS are:

1) Initialization:
    a. To select a set of neighborhood structures (NS): N_k, with k = 1, ..., k_{max}, and a initial solution x;
    b. To choose a stop criterion.
2) Iterations (repeat the following sequence until the stop criterion is reached):
    a. Let be $k \leftarrow 1$.
    b. Repeat until $k = k_{max}$, the following steps:
        i. Agitation. Randomly generate a solution x' of the k-th neighbor of x.
        ii. Local search (LS). To apply some local search method using x' as an initial solution; let x'' the local optimum obtained.
        iii. Replacement. If x'' is better than x, then $x \leftarrow x''$ and $k \leftarrow 1$; otherwise, let $k \leftarrow k + 1$;

The neighborhood structures (NS) used may be nested and staggered as they are at the reduced variable neighborhood search. The stop criterion may be, for instance, the maximum CPU time allowed, the maximum number of iterations, or the maximum number of iterations between two improvements. Note that the solution x' is randomly generated in the algorithm for avoiding premature loops, which may occur when other deterministic rules are used.

In order to construct the hybrid metaheuristic based on SA and VNS, we have started from a simple criterion: when the final temperature in SA reaches the maximum cooling phase, the solution obtained is used as initial solution for VNS. In this way, we expect a oscillating behavior at the first phase of the hybrid algorithm, meanwhile in the second phase only the best solutions of SA are obtained when the temperature is decreased.

Experimental results

The tests were randomly chosen, considering that the cost of the objective function depends of the geometrical compactness. However, the restriction of the bounds limits obtaining feasible solutions when the number of groups G is very large. As soon as G is small, the objective value is quite close to the optimum with a reasonable computational cost. But when the number of groups increases, the computational cost also increases and a penalization of the objective function value is evident (see Tables 1 and 2).

With respect to homogeneity, it is desirable that all the groups were made up of the same number of AGEBs (ideally the mean), however, in practice it is almost impossible. Let consider the case of 469 AGEBs clustered into 4 groups, then each groups should have 117.25 objects. In order to resolve this constraint, a variable named bounds was incorporated into the algorithm. This bound allows to increase or to reduce the number of objects conforming the groups. It is easy to deduce that these constraints lead to obtain a problem with a higher computational complexity. For experimental purposes, a set of non-systematical runs has been executed for evaluating the algorithm and its implications in its current state.

In Tables 1 and 2, *G* denotes the number of groups, OF is the cost of the objective function, IB is the inferior bound (the minimum number of objects), and SB is the superior bound (the maximum number of objects in the groups).Table 1shows a set of runs using the following parameters: RS: To ← 5000, Tf←.1, α← .98, L(t) ← 5 and VNS: NS← 2, LS ← 15.

In Table 2, it is shown the obtained results when the parameter NS is changed to 50, therefore, the complete parameters are established as follows: RS: To ← 5000, Tf←.1, α← .98, L(t)← 5 and VNS: NS← 50, LS← 15.

It is clear that the results are poor when the number of groups increases, existing a reduced balance on the number of groups. When the final result is a unfeasible solution, then the initial solution is given as output. On the other hand, good quality solutions are obtained with wide bounds. A random test for 35 groups with IB= 2 and SB= 38, and time of 9 seconds is presented. The second test for 2 groups with IB=220 and SB=236 was solved in 1 second. The parameters for both tests were the same used in the tests shown at Table 1.

Table 1. First set of tests: restricted bounds.

Tests	*G*	*OF*	*T (secs)*	*IB*	*SB*
1	3	31.0387	1	109	203
2	5	24.9812	0	66	122
3	7	23.4923	0	47	87
4	9	22.4401	1	36	68
5	11	20.1129	0	30	55
6	13	unfeasible	1	25	47
7	15	unfeasible	1	22	41
8	17	unfeasible	1	19	36
9	19	unfeasible	1	17	32
10	21	unfeasible	1	16	29
11	23	unfeasible	1	14	27
12	25	unfeasible	2	13	24
13	27	unfeasible	2	12	23
14	29	unfeasible	2	11	21
15	31	unfeasible	2	11	20

Table 2. Second set of tests: restricted bounds.

Tests	*G*	*OF*	*T (secs)*	*IB*	*SB*
1	3	30.5933	30	109	203
2	5	25.2013	16	66	122
3	7	22.319	22	47	87
4	9	19.2805	27	36	68
5	11	17.48	32	30	55
6	13	16.5917	37	25	47
7	15	15.7668	45	22	41
8	17	unfeasible	51	19	36
9	19	unfeasible	51	17	32
10	21	unfeasible	56	16	29
11	23	unfeasible	60	14	27
12	25	unfeasible	64	13	24
13	27	unfeasible	68	12	23
14	29	unfeasible	72	11	21
15	31	unfeasible	76	11	20

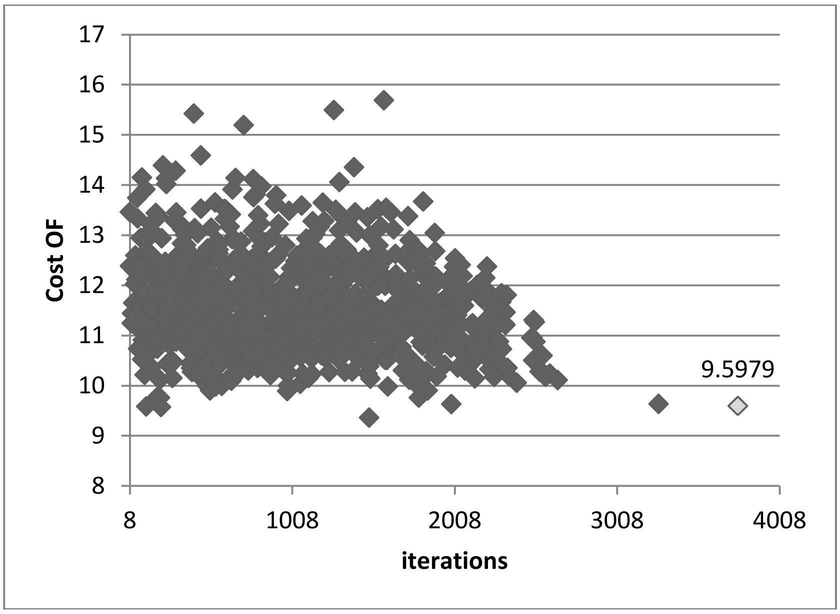

Fig. 1. Solutions for the 35 groups

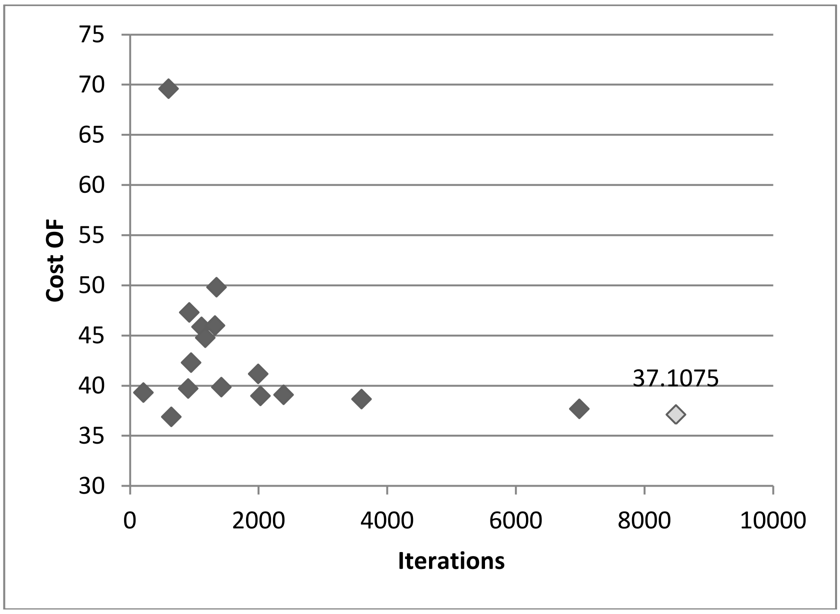

Fig. 2. Solutions for the 2 groups

The algorithm proposed obtained a good performance with the simple case. In Figures 1 and 2 it may be observed the oscillation of SA in a preliminary (first) phase, and the last 3 solutions correspond to VNS. This behavior means that the hybrid algorithm proposed allows obtaining the three best solutions.

Conclusions

In problems of territorial design, the solution of spatial restrictions are required. Sometimes it is sufficient with the solution of only one restriction (compactness, continuity, connectedness or homogeneity) but those are rare and very simple cases.

In this research work, we have incorporated homogeneity to the constraints of the compact grouping, however, the tests carried out reveal that the complexity increases when the bounds are close enough to the ideal average. The simpler case occurs when the groups are small (2 to 15 groups) with relaxed bounds. The harder case occurs when the number of groups are greater than 20 with close bounds.

So far we have compared our results with other methods, but we are working to obtain bound with Lagrangian relaxation, however we have not reported these results because we are analyzing the instances and also reviewing the method.

The remaining challenge is to solve the case of partitioning with a higher number of groups and solutions close to the average. For this purpose, we are constructing a multiobjective method based on the order theory, however in this case, the computational cost can be very high.

References

Bernábe B., Espinosa R., Ramírez J., Osorio M. (2011). A Statistical comparative analysis of Simulated Annealing and Variable Neighborhood Search for the Geographical Clustering Problem.Computación y Sistemas, 14 3: 295-308

Kalcsics, J., Nickel, S., Schröder, M. (2005). Toward a unified territorial design approach: Applications, algorithms, and GIS integration, TOP 13 (1) 1–56.

Kirkpatrick, CD, Gelatt, MP, Vecchi. (1983). Optimization by Simulated Annealing. SCIENCIE 220: 4598.

Kaufman, L., Rousseeuw, P. J. (1987). Clustering by means of medoids. Statistical Data Analysis based on the L1 Norm, North-Holland, Amsterdam 405-416.

Mladenović N, P. Hansen (1997). Variable neighborhood search. Computers & OperationsResearch, 24:1097-1100

Tavares P., Figueira F., Mousseau J., Roy V. (2007). Multiple criteria districting problems. The public transportation network pricing system of the Paris region. Annals of Operations Research, 154, pp. 69-9

Trejos Z. J., Castillo E.J., González V. (1990). Análisis multivariado de datos: Métodos y Aplicaciones Escuela de Matemática Universidad de Costa Rica (In Spanish).

Zoltners, A., Sinha, P. (1983). Towards a unified territory alignment: A review and model. Management Science, pp 1237-1256.