International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 6 (2014) pp. 45 - 54
© MIR Labs, www.mirlabs.net/ijcisim/index.html

A Bioinspired Proposal of Clustering Around Medoids with Variable Neighborhood Structures

María Beatríz Bernábe Loranca[1], Rogelio GonzálezVelázquez[2], Elías Olivares Benítez[3], Javier Ramírez Rodríguez[4] and Martín Estrada Analco[5]

[1,2,5] Benemérita Universidad Autónoma de Puebla BUAP, México, Puebla
beatriz.bernabe@gmail.com

[1,3]Universidad Popular Autónoma del Estado de Puebla México, Puebla
elias.olivares@upaep.mx

[4] Universidad Autónoma Metropolitana and LIA, Universitéd'Avignon et des Pays de Vaucluse, France
jararo@correo.azc.uam.mx

Abstract: **The artificial vision allows us to reduce a problem by means of techniques that have obeyed the study of the intelligence of living systems. A well-known technique is data mining and pattern recognition, which are disciplines dependent of artificial intelligence that from some data, allow the acquisition of knowledge and in particular, within data mining, a great application in the field of bioinformatics has been found.**

What is more, the big and diverse expansion of the amount of data produced by problems related to biological behavior has generated the necessity of constructing precise prediction and classification algorithms. The precision of classification algorithms can be affected by diverse factors, some of them considered generics in any automatic learning algorithm and, therefore, applicable to distinct research areas. These factors are the ones that have received attention in the field of automatic learning and pattern recognition, where different clustering algorithms are observed, in particular the automatic classification or better known as classification by partitions.

In this scenery, is important to discover an analogy between the way that some living beings form groups to survive in their environment finding an optimal sequence or structure or grouping their objects or belongings, and a classification by partitions algorithm.

The partitioning is an NP-hard problem, thus the incorporation of approximated methods is necessary. The heuristic that we expose here is Variable Neighborhood Search (VNS) focusing in the way that this heuristic does the search of neighbor conditions by means of neighborhoods to get a satisfactory solution, just like some living beings usually do it when they try to adapt to a neighborhood close to theirs or to the current space.

In this work, we focus on describing in a bioinspired way, a technique of data mining known as partitional grouping with the inclusion of VNS with the purpose of finding approximated solutions for a clustering problem.

Keywords: bioinformatics, clustering, data minning, partitioning, Variable Neighborhood Search

I. Introduction

Simulating the way some living beings survive, as well as the study of brain function, has been a challenge in some areas of exact sciences, these problems require a lot of effort but can cause a great impact generating useful techniquesto solve many problems.

In this point, bioinspired systems have arisen as a set of models that are based in the behavior and the way some biological systems act. These bioinspired systems can be seen as areas of Artificial Intelligence, Data Mining, and Operational Research, among others.For example, the application of hierarchical clustering to group stems with similar characteristics has been reported where experimental data indicate that agglomerative clustering can be successfully applied to the task of grouping stems in inflectional paradigms [1].

Many living beings, through different forms of searches that despite being intuitive and not systemic, solve a big amount of problems, however, most of the real problems require at least one search or grouping method to be solved [2, 3].

In this work we propose that observing the way and structure that most of living beings achieve to take care of searches when they group to accomplish a goal, is possible to setan analogy of the classification by partitions in a bioinspired way, that being a high computational complexity problem, it is found in the area of combinatorial optimization.

The bioinspired setting of the partitioning can be possibly stated as the partition that optimizes an objective finding the optimal structure and sequence of a minimum and efficient trajectory with the purpose of simplifying paths, routes and searches between its components with a minimum cost. A heuristic method within the partitioning is demanded, in this work we have proposed VNS due to the fact that it’s a trajectories algorithm, focusing in searches by environments-neighborhoods and neighborhood structures.

The article is organized as follows: Introduction as section 1, section 2 covers the nature of the Bioinspired aspects, section 3 describes social grouping to continue with section 4 that exposes partitioning. The section 5 introduces Neigborhood Search and section 6 presents VNS and partitioning. Finally in section 6 the conclusions and reflections on further work are discussed.

II. Bioinspired Aspects

The rapid grow of data, biological as well as process derived, has given place to propose methods for storing, organizing, classifying and efficiently managing information and also to the extract useful information from these data, being this last, one of the main challenges in computational biology.

These methods must provide a description beyond the data and the supplied knowledge in form of a bioinspired model.

Distinct biological domains exist where the Data Mining (DM) techniques are applied to the knowledge extraction. These problems have been categorized in genomic, proteomic, microarrays, systems biology, evolution and textmining [4].

The DM is the crucial phase of the Knowledge Discovery in Databases (KDD) process and consists in the development of computational algorithms that optimize an objective function or criterion using examples or previous experiences. The optimization criterion can be the accuracy of a determined model for a modeled problem, or the value of the evaluation function for an optimization one.

The optimization problem can be stated as the problem to find an optimal solution within a multiple solutions space. The selection of the optimization method is a crucial part to solve this kind of problem. The different optimization techniques for biological problems can be classified according to the type of solution found: exact or approximated methods. The optimization is a fundamental task in the modeled problems. In fact, the learning processes can be considered as the search for the best model that describes the data. In this search within the models space, any kind of heuristic can be employed. Therefore, the optimization methods can be considered as a part of the modeling.

In this point, even if the data mining that belongs to KDD is seen as a biological domain, then it may arise first, as a bioinformatic model and second, as computational algorithms that optimize a criterion using examples or previous experiences.

In this sense, is possible to propose that the partitioning being a subarea of DM and a problem of combinatorial optimization, the classification by partitions is in a bioinformatics context, and even of biological domain. Furthermore, the partitioning requires an approximated neighborhood search method and assuming that these searches have analogies with the way living beings examine neighboring conditions to adapt in their environment and that also organize in small groups, without getting apart and with short distance between them to achieve a goal, we can assume that in these conditions, the VNS partitioning is a bioinspired method.

We have as the main bioinspired algorithms the Neural Networks, Genetic Algorithms or Evolutionary Methods, Bee's hive and Ants' Colony.

However, it is possible to propose the partitional clustering together with VNS as a bioinspired method due to the fact that is based in the behavior of some beings living in community to organize by means of groups and thus find food or entities that in several occasions, such found entities also require being organized or ordered according to their properties.

In a conflict situation, to search objects of interest, usually living beings disperse in a random way in first place and when the object is reached, a call is made to the rest of the community to inform about the possible path, then the rest of the group, considers the path. But this is not enough; the least expensive path to the objects found is needed. This process is interpreted as an algorithm to the resolution of the minimum distance between two points. In this sense, well defined not bioinspired algorithms can be found in the computational literature, that are useful to solve the problem of the shortest path but very expensive in terms of computing time.

However, the difference between this algorithms and the partitioning-VNS that we propose, resides in that the firsts reach the solution by means of a serial and enumerative search (looking at a certain path until no further advance is possible) mean while the partitioning-VNS-bioinspired algorithm does it in a combined form.

Let's consider the objects as a community, then, first the number of groups is chosen with the focal objects (centroids), this can be interpreted that in order to organize, some communities usually choose a group representative that is capable of leading the other objects to form well defined groups.

By the search, they make the other objects come closer to them taking several paths until they achieve convergence by means of VNS and at the same time groups are created using the strategy of the final shortest accessible path, by which, all the objects community walks through. In terms of optimization, this search-trajectory is set as a cost function over the minimization of distances between objects.

To facilitate trajectories and searches is necessary to form groupsbased on the objective function that minimizes the distances between objects and centroids. This implies that we are in front of a problem that optimizes the minimum distance and that indicates, in topological terms, geometrical compactness.

An "optimal" solution for VNS partitioning is an approximated partition to a final compactness (global optimum). This approximated solution is the last solution that reaches the minimum value of the compactness objective function that has been accepted from all those "sub-optimal" or "local" solutions that are generated through the approximation process in VNS. Said "optimal" solution is the solution that reaches the minimum value of the compactness objective function.

With the incorporation of VNS in the partitioning, the heuristic process keeps going while the specific parameters of the local search and the neighborhood structures aren't fulfilled during the search process of the objective function [5].

III. Social Grouping

We have reviewed the social grouping in the psychology area with the goal of associating the behavior of grouped people to the grouping by partitions with VNS. Generally and in an informal way, a social group is understood as any meeting of people and the simple observable variables, of the set, of communication and specifically of the association, are justified. In these informal and primary terms, the essence of a group doesn't reside in the similarity or dissimilarity of its members but in its interdependence. A group can be characterized, as a "dynamic whole",this means that a change in the state of one of the parts modifies the state of any other part [6].

Group is as well a reunion of people that have personal and intimate relationships, which produces some psychological effects of changes in the behavior. The group has a dynamic action between the members, a common goal, relationship between size and function, will, consent and, capability to be guided by itself [7]. According to Natalio Kisneman a group is a set of individuals that interact with reference to a determined object (in the analogy with partitioning this is the centroid), the individual is pushed to belong to a group by several motivations of impersonal (metric) character, get friends, meet other personalities, need for security, acquiring knowledge, experiences, training [6]. Thus, a social group is understood as a set of people which members act and influence each other directly and the result is completely different of the individual action of each of the elements. The Social Psychology Manual affirms that in the group where different people work together in some tasks or that are present in a room at the same time, the important thing is not the *proximity*, but the *membership* to an official organization [7]. A group consists of a series of individuals that have relationships between them that makes them interdependent at a certain significant rank. In the group there isn't autonomy, it is supposed that the individualism is lost but not the freedom and consists of two or more people that share intrinsic values and that coordinate their behavior in such a way that allows them to act over those values (not empty groups with homogeneity over certain characteristics). By the middle of the XVII century the word group meant in french, a reunion of people, today is considered as personal relationships of 2 or more people that exchange opinions emitting a logical sense in a determined time. The term group can be understood as a sociological designation convenient to indicate any number of people, big or small, where such relationships have been established that only they can be imagined as a set. This kind of definitions, as many others in sociology and psychology almost always describe the relationships within a group that represent the characteristics of intimacy due to the fact that they are private in the people that form or take part in a determined group, furthermore they bring psychological effects, due to the intimate relationships, it produces changes in the behavior of the individual that loses its membership by virtue of the coexistence and affective pressures as a result of the interaction.The people acquire a certain form of mutual dependence (in partitioning similarity metric or interclass measure), leaving aside their individualist answers by convenience or necessity, giving origin to the interdependence, defined as the psychosocial expression that identifies the mutual necessity to achieve a common goal due to the fact that the people need each other to perform a task or in accordance to a function. The interaction plays an indispensable role in the interdependence effect, because of the ease of locating it in a determined space, place or site. In accordance to the above, and summing up, a social group consists of: 1) A set of two or more people in a determined space and defined time (not empty group), 3) They develop interactions, common goals and objectives generating a collective image (movement over an intra-class measure), 4) They generate functional interdependence and changes in their behavior where the results of the set are different to the individual ones (disjoint groups of empty intersection).

We characterize the group of people that reunite in a defined area and that directly influence each other generating new behaviors and emotional interdependence in this exchange that leads them to create goals and objectives almost common producing a specific image of the group which makes them easily identified by the rest.

The sources of satisfaction of the existing needs in the group cover as minimum, the following: a) attraction towards the members of the group: the proximity, the contact and interaction and physical attraction (similarity measure) b) relation with the objectives of the group (similarity measure and homogeneity characteristics). Within the social factors that motivate people to become part of a group; the status, the duty and activities of the group are distinguished in similarity (intra-class measure or similarity measure), the objectives of the group (cost function), need of membership, interests and similarities that differentiates them from other groups. The individuals that become part of a group (membership property), whose interactions become more frequent tend to develop their own environment, however, they can't detach from the external influence applied over them by the environment where the group moves over and if this one determines the norms and even the behavior, applies indirect influence until determining once again their structure, obviously depending on the wealth obtained from the interactions made. The clustering actsin this way, if the external influence of the centroid is better for an object, this becomes part of another group by the proximity or influence of the new centroid. According to the integration way the reformed groups are those which members know each other by being affectively united before building the group, the leader (centroid) is chosen by prestige and is the one in charge of assuring the permanence and the structure. The groups have membership; the members recognize each other (binary restriction in partitioning when an object must belong to one group only).

IV. Partitioning

A necessity exist of setting schemes for grouping to facilitate trajectories and searches, namely, proposing models and algorithms that solve the shortest path problem such that the objects being grouped must be very close to each other. This implies that we are in front of a problem that optimizes the minimum distance and indicates, in topological terms, geometrical compactness.

An "optimal" solution for partitioning-VNS is an approximated partition to a final compactness (global optimum). With the incorporation of VNS in the partitioning, the heuristic process keeps going while the specific parameters of local search and of the neighborhood structures aren't met during the reaching process of the objective function.

A. Partitioning Preliminaries

Among the most used techniques in data analysis (organization, structuration and grouping) are the classification techniques. Several are the classification methods that arise within data mining, but our interest focuses in grouping by partitions that produce asa final solution a unique cluster partition of not overlapped objects in a particular number of groups (k) previously specified, as a result of the minimization or maximization of an objective function. Regularly, these methods start with an initial partition of the set of objects into k clusters, a centroid is defined for each of them; then, each object is located in the cluster that has its closest centroid and subsequently, calculates the new centroids to reallocate each object again. So on and so forth until no changes are produced in the clusters or a cost function is reached. The most important constraints consist in that each group has at least one element and each element belongs to only one group.

In partitioning, generally the objects are represented by D descriptive attributes in form of vectors in the space R^D, and with a similarity comparison measure, such as the distance, the clusters are created with similar objects. In the group formation process there is no previous knowledge about how a cluster must be formed; for that reason, the clustering process is also known as unsupervised classification, relevant method in DM.

In the grouping the information of a series of variables for each object is used and according to these variables, the similarity between these objects is measured. Once the similarity is determined, the objects are grouped in internally homogeneous groups and different between them.

On the other hand, the dissimilarity is important, contrary to the similarity, this measures how different two individuals are (namely, is more probable that it belongs to the same class or group) and as it gets higher, the more different they will be. For better accuracy, the distance expresses proximity or remoteness between two objects. In math, the distance between two points of the Euclidean space equals the longitude of the straight line segment that unites them.

B. Mathematical Expression of Partitioning

In the classification by partitions we have $\Omega = \{x_1,..,x_n\}$ the finite set of *n* objects to classify; $k < n$ the number of clusters where the objects are desired to be classified. A partition $P = \{C_1,..,C_k\}$ of Ω in k clusters $C_1,..,C_k$ is characterized by the following conditions:

1. $\Omega = \bigcup_{i=1}^{k} C_i$
2. $C_i \cap C_j = \varnothing$, for each $i \neq j$

C. Description of the clustering problem as an optimization one

Suppose that $x_i \in R^D$, and that k is an integer number previously known, the clustering problem consist in finding a partition P of Ω such that each cluster is formed under a similarity metric evaluated by means of the Euclidian Distance which domains a setof attributes D such that $f : R^D \to R$. To measure the distance between two objects from Ω the formula (1) is used $X_i = (x_{i1},..,x_{iD})$ and $X_j = (x_{j1},..,x_{jD})$ is

$$d(X_i, X_j) = \sqrt{\sum_{l=1}^{D}(x_{il} - x_{jl})^2} \qquad (1)$$

The objects of a cluster are similar when the distance between them is minimal; this allows formulating the objective function*f*, as:

$$\sum_{j=1}^{k} \sum_{x_i \in C_j} d(x_i, \overline{x_j})^2 \qquad (2)$$

This is, minimizing (2) is desired; where x_j known as representative element of the cluster, is the measure of the elements of the cluster C_j

$$x_j = \frac{1}{|C_j|}\sum x_i \in C_j \qquad (3)$$

and belongs to the center of the cluster. Under these characteristics, the clustering is a combinatorial optimization problem, and has been proven that is an NP-hard one [8].

V. Neighborhood Search

The procedures of *Neighborhood Search* NS run over the space of solutions *U* by means of a set of transformations or moves. The solutions that are obtained from another through one of the possible moves are known as the neighbors of this solution and constitute its *neighborhood*. The set of possible moves gives place to a neighborhood relationship and a *neighborhood structure* in the solutions space. The general scheme of a neighborhood search procedure consists in generating an initial solution and, until a stopping criterion is met, iteratively selecting a move to modify the solution.

The neighborhood of a solution is formed by the solutions that can be accessed from it by one of the possible moves.

Formally, a *neighborhood structure* over a space or search universe *U* is a function E: $U \to 2^S$ that associates to each solution x ò U a neighborhood $E(x) \subseteq U$ of neighbor solutions to *x*. A big amount of heuristic methods proposed in the literature belong to the class of neighborhood search procedures [9].

The description in pseudocode of the neighborhood search is shown below.

```
Procedure neighborhood search
{
x ← generate Solution (U)
x* ← x
Do {
      x ← select solution (E(x));
        If (object (x) improves object (x*))
x* ← x;
        }while (not stop criterion)
}
```

The election of the neighborhood structure is fundamental in the success of the search procedures since it determines the quality of the set of moves applied. The combined moves appear when several moves are executed subsequently over a solution. An adequate combination of moves enriches the neighborhoods, which allows taking wider steps in the approaching to the optimum. An important characteristic of the moves is the feasibility of the contributed solutions.

Formally, the procedures that only take into account feasible moves are associated to the concept, somewhat more restrictive, of neighborhood structure as a function E: $S \rightarrow 2^S$ that associates to each feasible solution $x \epsilon S$ a neighborhood $E(x) \subset S$ of feasible solutions to x.

The main neighborhood search metaheuristics focus only on the selection procedure of the move. However, besides the selection of the neighborhood structure where the search will be articulated on, there are other relevant questions in the success of the neighborhood search procedure, such as the evaluation of the objective function, the procedure of generating the initial solution and the stopping criterion.

A. *Variable Neighborhood Search*

The variable neighborhood search (VNS) is a recent metaheuristic that consist in changing in a systematic way the neighborhood structure [9, 10]. The original idea was to consider distinct neighborhood structures and changing them systematically to escape from local minimums. The basic VNS obtains a solution from the neighborhood of the current solution, executes a monotonous local search LS from it until reaching a local optimum, that replaces the current solution if there's been an improvement and modifies the neighborhood structure otherwise.

The representations of the solutions for big problems may be difficult to read, and many problems, the Euclidean problems in particular (such as partitioning), allow a natural decomposition, a *focus* routine allows the representation of the previously said information to selected subproblems; namely, in some region of the space in which the route is traced.

B. *Fundamental Schemes*

An optimization problem consists in finding, within a set of *X*feasible solutions, the one that optimizes a function *f(x)*. If it is a minimization problem, it's formulated as follows:

Min $\{ f(x) \mid x \in X \}$ (1)

Where *x* represents an alternative *solution*,*f*is the *objective function* and *X* is the *space of feasible solutions* of the problem.

An *optimal solution* x^* (or global minimum) of the problem is a feasible solution where the minimum of (1) is reached.A neighborhood structure in the space of solutions *X* is an application $N : X \rightarrow 2^x$ that associates to each solution $x \epsilon X$ a neighborhood of solutions $N(x) \subset X$, which are called neighbors of *x*.

A solution $x^* \grave{\partial} X$ is a *global minimum* of the problem (1) if there is no solution $x \grave{\partial} X$ such that $f(x) < f(x^*)$. We say that $x' \grave{\partial} X$ is a local minimum with respect to N_k if there is no solution $x \grave{\partial} N_k(x') \subseteq X$ such that $f(x) < f(x')$.

The VNS is based in three simple facts:

1. A local minimum with a neighborhood structure it's not necessarily one with another structure.
2. A global minimum is a local minimum with every possible neighborhood structure.
3. For several problems, the local minimums with the same or different neighborhood structure are relatively close.

This last observation, that is empirical, implies that the local optimums provide information about the global optimum.

The facts 1 to 3 suggest the employment of several neighborhood structures in the local searches to tackle an optimization problem. The change of neighborhood structure can be done in a deterministic, stochastic or deterministic and stochastic way at the same time [10].

In summary, for VNS is necessary to consider distinct neighborhood structures and systematically change them to escape from the local optimums. The basic VNS obtains one solution from the neighborhood of the current solution, executes a monotonous local search from it until a local optimum is reached, that replaces the current solution if an improvement has been made and then the neighborhood structure is change otherwise. The most useful variant in our algorithm is the variable neighborhood search descend (VND) that applies a monotonous neighborhood search systematically changing the neighborhood structure every time a local minimum is reached.

The figure 1 shows graphically the behavior of VNS (similar a VND) for the algorithm we have proposed. Each solution *f(x)* is a partition that is built in a neighborhood N(x) and with local search is possible to find another better, in such a way that the algorithm keeps iterating in this neighborhood until it finds a worse solution to move to another neighborhood, or it moves to another neighborhood and creates another solution with a stop condition of local search.

This process is repeated until N(x) meets a given number of neighborhood structures.

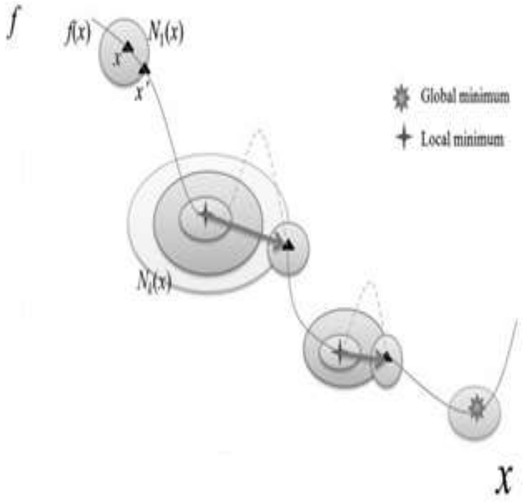


Figure 1. Graphic for VNS

VI. Variable Neighborhood Search and Partitioning

Once we have collected different theoretical aspects of clustering in data mining, the social grouping in psychology, clustering in general and VNS, we have integrated VNS with partitioning in an algorithm to achieve a bioinspired proposal of partitioning to solve distinct kind of grouping problems, especially those problems that involve objects with coordinates in R^2.

Before presenting the algorithm VNS-partitioning we expose an analogy of clustering with VNS:
People (individuals) ↔ minimal units or objects of clustering,
Partition ↔ Disjoint groups of people (individuals) ↔ a clustering solution in VNS,
Cluster ↔ group of people ↔ components of a solution.
Due to the fact that the formation of social groups of people is stimulated by the necessity that the individuals have of self-development and surviving as optimality conditions, it is required the acceptance of a set of norms to belong and to remain in it, this is, restrictions. Thus the elements of a combinatory optimization model can be distinguished, where the solutions space is formed by subsets of the natural numbers.

The concept of partitioning with VNS: a bioinspired approach is related to the generation of groups of objects emulating the way people group into social groups. Assume a group of people trying to look among the crowd for similar people. The application consists of the following:

1. Once an initial solution has been obtained it is expected that the groups are not necessarily at this point compact nor connected, especially if the initial solution is randomly generated.
2. The reallocation of individuals starts as they try to look for colleagues that are similar to them, this is close to them under certain criteria or metric, this is, similar individuals are grouped following a leader (centroid or group representative).
3. In each iteration, an individual reallocation is produced. Each individual starts to evaluate under the given metric, if it belongs to the group it should. This happens until obtaining an improved configuration.
4. As stopping criteria we can put that the distance between two consecutive configurations is below a given epsilon.
5. The process is repeated until the configuration of the partition has a dissimilarity rank between the groups, this will produce compact and connected groups.

The groups must have more than one individual due to the fact that the proposed restrictions and the metric are in reality to a partitioning model.

However the efforts to propose a bioinspired VNS analogy with the group behavior of living beings aren't enough, it is needed to find an accurate analogy of ´some nomad or ethnic group.

A. *VNS Metaheuristic*

VNS metaheuristic, is based on the observation that local minima tend to cluster in one or more areas of the searching space. Therefore when a local optimum is found, one can get advantage of the information it contains. For example, the value of several variables may be equal or close to their values at the global optimum. Looking for better solutions, VNS starts exploring, first the nearby neighborhoods of its current solution, and gradually the more distant ones. There is a current solution S_a and a neighborhood of order *k* associated to each iteration of VNS. Two steps are executed in every iteration: first, the generation of a neighbor solution of S_a, named $S_p \in N_k(S_a)$, and second, the application of a local search procedure on S_p, that leads to a new solution Sol. If Sol improves the current solution S_a, then the searching procedure will restart now from *sol* using *k=1*. Otherwise, *k* is incremented and the procedure is repeated from S_a. The algorithm stops after a certain number of times that the complete exploration sequence $N_1;N_2; \ldots ;N_{kmax}$ is performed. With the inclusion of the clustering schemes and the general procedure for VNS, a custom algorithm was obtained and we have named it neighborhood structures in partitioning.

B. *Partitioning Algorithm with neighborhood structures*

```
INPUT: Number of groups of objects corresponding to the K centroids, parameter
values for VNS and the distances matrix.
OUTPUT: the objects belonging to each group, the parameters values, the initial and final
execution time and number of iterations, the iteration number associated with the best
value of the objective function.

Let n be the number of objects to classify
$U_g(i,j)$ denotes that the object i is assigned to the centroid j for i=1,...,n; j=1,…,k
Let M={$M_1$,$M_2$,…,$M_k$} be a solution of K centroids
MaxVNS /* maximum number of iterations to go over all the neighborhood search */
MaxLS /* number of iterations of Local Search (LS) for each neighborhood */

1. Generate initial random centroids:
M = {$M_1$, $M_2$,…, $M_k$}
BEGIN
Current _cost← Cost (M)
 WHILE cont<MaxVNS DO
  BEGIN
k-neighborhood← 1
      WHILE k-neighborhood <> n DO
    BEGIN
C ← Generates a random solution with a k-neighborhood
  Sol_neighborhood←LocalSearch (C);
   IF
(Cost(Sol_neighborhood)<current_cost)
THEN
M ←Sol_neighborhood;
ELSE
k-neighborhood←k-neighborhood +1;
ENDIF
 END WHILE
END
 END WHILE
Cont← cont+1
END

2. Cost Function (Sol)
/* Determine the quality of the solution Sol, how much the objective is minimized */
BEGIN
i← 1 /* Initialize the first object */
cost← 0
 WHILE (i≤ n) DO
  BEGIN
/* For each object in Ug do */
   IF ($Ug_i$is not a centroid) THEN
    BEGIN
```

```
dmin←dist($Sol_1$ , $Ug_i$)
/* Represents the distance between the object and the $Sol_1$ (first centroid where Sol represents the set of centroids). The distance between each object and its nearest centroid is calculated */
j← 2
/* Go to the second centroid  */
  WHILE (j ≤ k) THEN
      BEGIN
   IF (dist ($Sol_j$,$Ug_i$) <dmin) THEN
/*Calculate the distance between the object i and the  $Sol_j$(another centroid) */
  dmin←dist ($Sol_j$, $Ug_i$)
   END IF
  j←j + 1
/* Go to the next centroid  */
  END WHILE
  cost←cost + dmin
   END IF
i←i + 1
 END WHILE
Cost(Sol) ← cost
END
```

The figure 2 represents how a partitioning solution would be in a graph for our problem.

An example of graph partitioning has been reported in [11], where the clustering has built partitions for a case of the mobility of the Dengue Mosquito.

In accordance to the algorithm 1 we have built and considering the figure 1, each f(x) in N(x) is a partitioning solution as it is shown in figure 2.

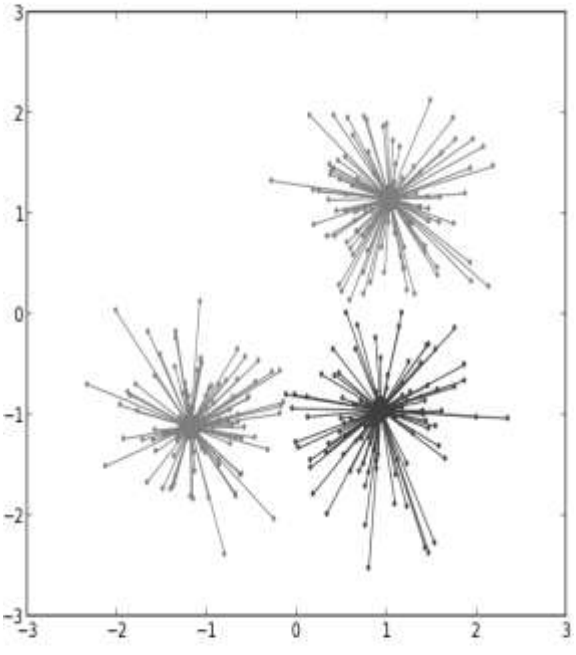


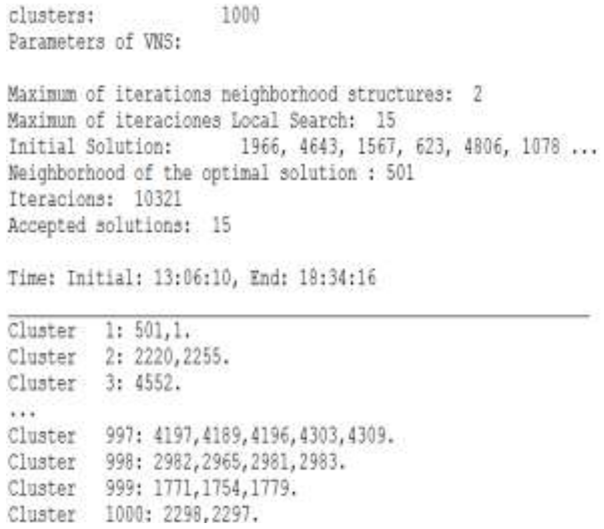
Figure 2. Clustering example

In this work, the grouped data is geographical and we haven’t given the results in graphs due to the fact that a graphical interface with a Geographic Information System GIS has been implemented [12], in such a way that it has been possible to show the results in maps as we will soon see.

The proposed algorithm can be applied to diverse clustering problems and different types of data as long as this has a position in R^2. The data that is processed in our algorithm must have a spatial position, like any individual.

Living beings represent a case of study within bioinspired clustering algorithms due to the fact that most of its organization and administration problems require of systematic processes to group, searching for near neighbors.

If we talk about approaching, grouping and organizing, the bioinspired algorithms of animals like bee swarms, mosquito swarm, ants, etc. have deserved special attention. However fixed points in the plane can also be processed by the algorithm we have proposed. In this sense, we have considered 4965 geographic data from a census process (Agebs: basic geostatistical areas). The data belongs to the Metropolitan Zone of Mexico Valley (ZMVM) [13]. The application is a population sampling where 1000 groups are required.The parameters for VNS are LS (local search)=2 and (neighborhood structures) NS=15. In Figure 3 the obtained result can be observed.



```
clusters:            1000
Parameters of VNS:

Maximum of iterations neighborhood structures:  2
Maximun of iteraciones Local Search:  15
Initial Solution:      1966, 4643, 1567, 623, 4806, 1078 ...
Neighborhood of the optimal solution : 501
Iteracions:  10321
Accepted solutions:  15

Time: Initial: 13:06:10, End: 18:34:16
________________________________________________________________
Cluster   1: 501,1.
Cluster   2: 2220,2255.
Cluster   3: 4552.
...
Cluster   997: 4197,4189,4196,4303,4309.
Cluster   998: 2982,2965,2981,2983.
Cluster   999: 1771,1754,1779.
Cluster   1000: 2298,2297.
```

Figure 3. Results VNS for 1000 groups of ZMVM

The figure 4 shows the 15 solutions accepted by VNS.

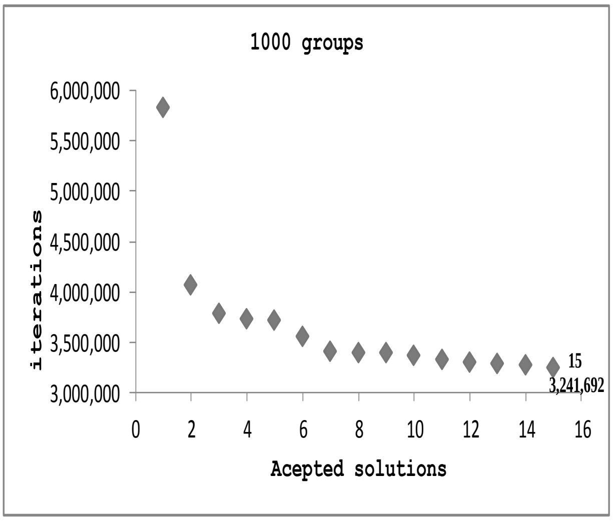


Figure 4. Iterations VNS and Objetive Cost

The figure 5 represents the associated map to the solutions of the MZVM for 4965 geographical units grouped with bioinspired VNS with the parameters previously mentioned. The map has been obtained exploiting the tools for developers of MapX in such a way that software has been implemented to create maps of diverse metropolitan zones with a specific format in the input data [12, 14].

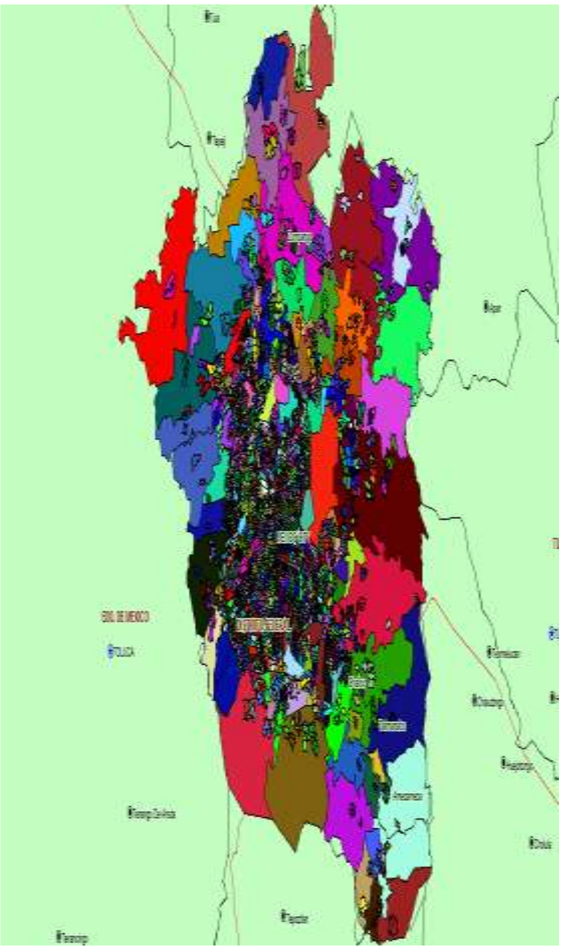


Figure 5. Solution for 1000 clusters for ZMVM whit MapX

C. *Comparison of results (Metropolitan Zone of Toluca Valley ZMVT)*

We have taken a representative example: to evaluate the efficiency of the proposed algorithm, we obtained a “best solution”with an exact algorithm (Partitioning Around Medoids, PAM) [15]. We compare the results for ZMVT (469 objects). The quality of the obtained solutions is measured with the relative duality gap, defined as gap=((Sol$_{PAM}$-Sol$_{VNS}$)/Sol$_{VNS}$)100, where Sol$_{PAM}$ denotes the optimal solution of the problem (best solution) and Sol$_{VNS}$ denotes the solution obtained by descendent VNS included in the algorithm that we have proposed.

At first 4 corresponding instances were defined with 8, 12, 18 and 24 groups for the comparison as shown in Table I. For these instances the average gap is 8% with regard to VNS and where Sol$_{SA}$ denotes the obtained solution with Simulated Annealing (SA) and is shown that VNS offers better solutions than simulated annealing [16].

Table 1. Solutions obtained by PAM, VNS and RS for ZMVT

Group	Sol$_{best}$	Sol$_{VNS}$	Sol$_{SA}$
8	17.432	17.691	18.946
12	14.120	15.159	15.635
18	11.098	12.441	12.09
24	9.279	0.8371	11.240

On the other hand, a total of 32 instances of the problem were generated for the next values of *n* or size of the matrix: 100, 200, 300, 400, 450, 500, 600 and 700. For each of these values of *n*, four values of groups were considered: 8, 12, 18 and 24. The matrices with size under 469 were taken as sub-matrices of ZMVT and the ones with size is over 469 were completed generating random numbers in the range (0, 1) obtaining the optimal values with PAM and compared to VNS there is a12% gap. In accordance to the obtained results we can say that our methodology has moderate efficiency, but it is better than SA in some cases.

As second example we have taken the data of the ZMVT for small instances of 8, 16, 32, 48 and 100 groups, with a reasonable computing time between 15 and 25 seconds for 15 iterations in neighborhood structures and 2 in local search (figure 6).

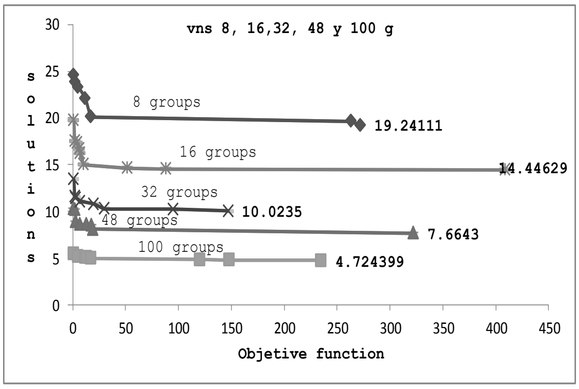


Figure 6. Solutions for100 clusters

The figure reflects the expected and trivial behavior in clustering:the bigger the number of groups the smaller the cost function.

VII. Conclusions

We have proposed a VNS-partitioning as a bionsipired method due to the fact that VNS finds solutions in environments and structures (as some living beings usually do).

Readings about social behavior in the field of psychology and bioinspired algorithms and optimization for knowledge discovery [7, 17], have been needed for the conclusion of this work, however, it still remains the precision and formalization of the analogies that we pursue. In this point, in the literature we have found the text known as the Argonauts of the occidental pacific, which is a classic work of anthropology, where the author (Malinowski) describes the culture of the habitants of the archipelago [18]. The axis of the study is the analysis of the kula, ceremonial exchange system of necklaces and bracelets, several of them with a great traditional value, with its own name and stories relative to the people that have possessed them, (here it has been focused our attention to assume that mechanism of exchange is similar to the partitioning over Medoids when in the second phase of the algorithm exists the exchange between objects and centroids).

Only the most influential characters of the islands take part in the kula, which is accompanied by rituals and ceremonies. This ceremonial exchange is the linkage point employed by the anthropologist to explain the customs, myths, traditions, technologies and social hierarchy of the trobriands, and to prove that this rite involves other institutions like magic, the myths and the social ranks where the author catalogues the kula as a reciprocity mechanism which function is to reinforce the relationships between the members of the different tribes of the place, to temper the conflicts between them and thus facilitate other kinds of relationships, in particular the commercial ones and to group without conflicts. In this point, we have placed a special interest in the conduct of the kula to comprehend the way in that they get socially organized in accordance to their culture and the exchanges that characterize them and to be able to establish an analogy with an ethnic culture and the clustering within data mining, statistics and optimization.

Finally, partitioning is a grouping process required in many tasks of living beings.To group objects, the searches in the partitioning form clusters with representative objects. The results of these groupings leave information to let the rest of the objects get close to the representative (centroid) and thus find a better solution. The combination of these two methods promises to obtain good quality clusters with a low computational cost.

Acknowledgment

The first author acknowledges support from CONACyT project extension of postdoctoral 2012-2013 for development this work.

References

[1] Z. Detorakis, G. Tambouratzis. "Clustering Techniques for Establishing Inflectionally Similar Groups of Stems",*International Journal of Computer Information Systems and Industrial Management Applications*, ISSN 2150-7988 (4) pp. 219-227, 2012.

[2] S. F. Uribe Tarbodas. "Reseña de los argonautas del Pacifico Occidental", *Boletin de Antropologia*, Universidad de Antioquia: Medellin Colombia, 19 (36), pp. 394-400, 2005.

[3] O. Kramer.Self-Adaptive Heuristics for Evolutionary Computation, *Springer-Verlag Berlin Heidelberg*, 2008.

[4] P. Larrañaga, B. Calvo, R. Santana, C. Bielza, J. Galdiano, I. Inza, J. A. Lozano, R. Armañanzas, G. Santafé, A. Pérez and V. Robles. "Machine learning in bioinformatics", *Briefings in Bioinformatics*, (7), pp. 86-112, 2006.

[5] B. Bernábe, J. Ramírez, J. Espinosa, M. A Osorio, R. Aceves."An Adjusted Variable Neighborhood Search Algorithm applied to the Geographical Clustering Problem",*Research in Computer Science*, ISSN 1870-4069, (42), pp. 113-125, 2009.

[6] C. E. Torres. "Grupos, teorías y experiencias académicas", unpublished.

[7] M. Hunt. "Personality and the behavior disorders a handbook based on experimental and clinical research",*Edited by J. Hunt Director, Institute of Eelfare Research, Community Service Society of New York*, (2), pp. 692-714, 1944.

[8] E. Vicente., L. Rivera., D. Mauricio. "Grasp en la resolución del problema de clustering", ISSN 1815-0268, (2), pp. 16-25, 2005.

[9] A. D. Pelta. "Algoritmos heurísticos en bioinformática". Tesis doctoral, *Departamento de Ciencias de la Computación e Inteligencia Artificial. Universidad de granada*, 2000.

[10] P. Hansen, N. Mladenovic, J. Moreno Pérez. "Variable Neighbourhood Search", *Revista Iberoamericana de Inteligencia Artificial*, ISSN 1137-3601, (19), pp. 77-92, 2003.

[11] M. Bernábe, M. Rodríguez, R. Martínez , J. Ramos and E. O. Benitez."Adaptation of a Clustering Algorithm and Mosquito Swarm to a problem of ovitraps for the Dengue Mosquito Vector*", Fourth World Congress on Nature and Biologically Inspired Computing*, 2012.

[12] E. Zamora, "Implementación de un Algoritmo Compacto y Homogéneo para la Clasificación de AGEBs bajo una Interfaz Gráfica". Tesis de Ingeniería en Ciencias de la Computación, *Benemérita Universidad Autónoma de Puebla*, México, 18-27, 2006.

[13] INEGI (s.f.). Retrieved from http://www.inegi.org.mx/est/contenidos/espanol/soc/sis/microdatos/default.aspx

[14] MapX Developer´s guide. MapInfo corporation. Troy, NY.

[15] L. Kaufman, P. J. Rousseeuw. "Clustering by means of medoids", *Statistical Data Analysis based on the L1 Norm, North-Holland*, Amsterdam, pp. 405-416, 1987.

[16] B. Bernábe, J. Espinosa, J. Ramírez J., and M. A Osorio, Statistical comparative analysis of Simulated Annealing and Variable Neighborhood Search for the Geographical Clustering Problem, Computación y Sistemas, vol. 42, no. 3, 2009, pp. 295-308.

[17] R. Sarker, H. Abbass, C. Newton. "Heuristics and Optimization for Knowledge Discovery",*New South Wales, Australia Idea Group Publishing*, pp. 208-225, 2002.

[18] A. V. Malinowski."Argonauts of the Western Pacific", *Edicions 62, S. A., 1973, Editorial Planeta-De Agostini*, S. A. Barcelona (España), ISBN 84-395-0140-4, 1986.

Author Biographies

María Beatríz Bernábe Loranca was born in the city of Puebla, Mexico. He received the B.S. degree in Computer Science from Benemérita Universidad Autónoma de Puebla (BUAP), Mexico. In 2010, she received the Doctorate degree in Operations Research from the Universidad Nacional Autónoma de México (UNAM). Since 1995, she has been a professor at the School of Computer Science of BUAP, where she works in databases and statistics. She belongs to the National System of Researchers with Level Candidate (SNI). Her research interests are: combinatorial optimization, territorial design and multiobjective techniques.

Rogelio González Velazquez received the Master degree in Operations Research on Universidad Nacional Autónoma de México in 2000 and Ph. D. degree in Logistic and Supply Chain on Universidad Popular Autónoma de Puebla (Mexico) in 2012. Currently, he is a Professor-Researcher of the Faculty of Computer Science of the Benemérita Universidad Autónoma de Puebla (Mexico). He is interested in research about operations research, combinatorial optimization, metaheuristics and particularly location-allocation problems.

Elias Olivares Benítez Professor in the areas of Supply Chain Management, Industrial Engineering, and Materials Engineering. His research interests are in modeling and optimization of logistics, manufacturing, supply chain, and service systems, with multiple objectives. He has used Scatter Search, Tabu Search, GRASP, CPLEX, and hybridizations. He is member of the Institute for Operations Research and the Management Sciences, the Mexican Society for Operations Research, and the Mexican Association of Logistics and Supply Chain

Javier Ramírez Rodríguez is Full Professor in the Departmento de Sistemas at Universidad Autónoma Metropolitana in Mexico City. He is Doctor in Mathematics from Universidad Complutense de Madrid. Visiting Professor at LIA Universitéd'Avignonet des Pays de Vaucluse, France. His research interests are in heuristics methods, network optimization and soft computing. He belongs to the National System of Researchers Level I (SNI I).

Martín Estrada Analco received the Bachelor degree in Mathematics and Master degree in Mexican Letters 1987 and 2002 respectively on the Benemérita Universidad Autónoma de Puebla (Mexico). Currently, he is a Professor-Researcher in the Faculty of Computer Science for the Benemérita Universidad Autónoma de Puebla (Mexico). He is interested in research about multivariable statistic and operations research.