Nature-inspired components of the Scatter Search[1]

David Beltrán Cano, Julio Brito Santana, Clara Campos Rodríguez, Ignacio J. García del Amo, Miguel García Torres, Francisco Javier Martínez García, Belén Melián Batista, José A. Moreno Pérez, J. Marcos Moreno Vega, and Raquel Rivero Martín

GCI – Intelligent Computation Group,
Dep. Estadística, I.O. y Computación
Instituto Universitario de Desarrollo Regional
Universidad de La Laguna, 38271 La Laguna Spain

Phone: +34-922318186, Fax: +34-922318170
email:{dbeltran,jbrito,ccampos,igdelamo,mgarciat,
fmartine,mbmelian,jamoreno,jmmoreno,rrivero}@ull.es

ABSTRACT: The *Scatter Search* is an evolutionary algorithm in which a moderate sized set of solutions or individuals evolves due to mechanisms of combination between individuals. Unlike other strategies of operation rules usual in genetic algorithms, the search for a local optimum in a Scatter Search is guided. Scatter Search consists of five components processes: 1) *Diversification Generation Method*, that generates sets of diverse individuals, 2) *Improvement Method*, that improves a individual to transform it in a better one, 3) *Reference Set Update Method*, which builds and updates the reference set consisting of good and diverse individuals, 4) *Subset Generation Method*, to produce subsets of individuals of the reference set to be combined, and 5) *Individual Combination Method*, that combines the individuals in the produced subsets. In this paper we analyse the main characteristics of these components from their nature-inspired point of view.

KEYWORDS: Scatter Search, Nature-inspired components,

1. INTRODUCTION

In a Scatter Search [14], from a wide population of solutions or individuals, a moderate sized set of individuals is considered. The reference set is generated and updated attempting to intensify and diversify this evolving set. After combining several individuals in the reference set, a local search procedure is applied to the resulting individuals, and the reference set is updated to incorporate both good and diverse individuals. These steps are repeated until a stopping condition is met. Thus the reference set evolves by a combination procedure that uses its elements to obtain others exploiting the knowledge of the problem at hand. Unlike the population in Genetic Algorithms ([12], [13]), the evolving set of individual in Scatter Search, the reference set, is relatively small. The principles of the Scatter Search metaheuristic were first introduced by Fred Glover in the 1970s as an extension of formulations for combining decision rules and problem constraints. This initial proposal generates new rules taking account of characteristics in several parts of the solution space [6]. An important feature of Scatter Search is its association with the Tabu Search metaheuristic and the fact that the search can be improved by including particular forms of adaptive memory and associated memory-exploiting mechanisms. Scatter Search has an implicit form of memory, which can be considered as an inheritance memory, since it keeps track of the best individuals found during the search and selects their good features to create new ones. Since this inheritance memory could be not sufficient, the adaptive memory principles of Tabu Search can improve the effectiveness of Scatter Search. Consequently, the Scatter-Tabu Search Hybrid has been widely applied to solve optimization problems. The Scatter Search Template, proposed by Glover in 1998 [8], summarizes the general description of Scatter Search given in [7].

[1] This research has been partially supported by the projects TIC2002-04242-C03-01 (70\% of which are FEDER funds) and PI042004/088}

The basic Scatter Search includes five main components. The procedure stars generating a large set of diverse individuals, *Pop*, which is obtained using the first component, the *Diversification Generation Method*. This procedure creates an initial population, which must be a wide set consisting in diverse and good individuals. Several nature-inspired strategies can be applied to get a population with these properties. A set of good representative individuals of the population is chosen to generate the reference set. The good individuals are not limited to those with the best objective function values. The individuals considered to be included are those with the best objective function values and also the individuals that are most diverse with the individuals already included. Several subsets of individuals from the reference ser are then selected by the second component; the *Subset Generation Method*. The third component of the Scatter Search is the *Individual Combination Method* that combines the individuals in each subset taking account their good characteristics. Usual implementations consider a subset generation method that chooses couples of individuals and the combination is a guided task that selects characteristics of those individuals similar to the nature-inspired crossover operations. Then, the fourth component, the *Improvement Method* is applied to the results of the combination to get new improved individuals. This method is a guided task based on transformations similar to the nature-inspired mutation operation. Finally, the fifth and last component, the *Reference Set Update Method,* uses the obtained individuals to update the reference set using criteria based on both, quality and diversity. A comprehensive description of the elements of Scatter Search proposed by F. Glover can be found in [8], [9], [10] and [11].

The rest of the paper is organized as follows. Next section describes how the scatter search works. Then we describe the main characteristics of the implementations of the components of the Scatter Search for standard well-known problems. The paper ends with the conclusions and references

2. SCATTER SEARCH

Scatter Search [14] is a population-based metaheuristic that uses a reference set to combine its individuals and obtain others. The method generates a reference set from a wider population of individuals. Then a subset is selected from this reference set. The selected individuals are combined to get new individuals to be used in an improvement procedure. The individuals resulting from these improvements can motivate the updating of the reference set and even the updating of the population. The basic Scatter Search procedure is shown in figure 1.

```
procedure Sequential Scatter Search

begin
      repeat
            Create Population;
            Generate Reference Set;
            repeat
                  repeat
                        Select Subsets;
                        Combination Subsets;
                        Improvement Combined;
                  until (StoppingCriterion1);
                  Update Reference Set;
            until (StoppingCriterion2);
      until (StoppingCriterion3);
end.
```

Figure 1: Scatter Search procedure.

The initial population must be a wide set of disperse individuals. However, it must also include good individuals. Several strategies can be applied to get a population with these properties. The individuals for the population can be created, for instance, by using a random procedure to achieve a certain level of diversity. Then a simple improvement heuristic method must be applied to these individuals in order to get better ones. The initial population can also be obtained by a method that provides at the same time disperse and good individuals like the well known GRASP constructive methods [17] and [2].

A set of good representative individuals of the population is chosen to generate the reference set. These good individuals are not limited to those with the best objective values. By good representative individuals we mean individuals with the best objective values as well as disperse individuals. Disperse individuals should reach different local optima by the improvement method. Indeed, an individual may be added to the reference set if the diversity of the reference set improves. So the reference set must consist of a set of disperse and good individuals selected from the populations. The criteria for updating the reference set, when necessary, must be based on measures of the quality of the individuals and measures of diversity between the new individuals and the existing individuals.

One or several subsets of individuals from the reference set are selected for applications of a combination method to individual subset to get good others individuals to be improved by an improvement procedure. In general, the selection method consists in choosing all the subsets of a fixed size. The combination method tries to combine good characteristics of the selected individuals to get new current individuals. The purpose is to get good individuals which are not similar to those already in the reference set. The possible improvement methods applied to the individuals range from simplest local searches to a very specialized method. A very simple procedure is a local search based on basic transformations (mutations or moves) consisting of iteratively selecting the best improving move or a first found improving move. The procedure must allow the use of tools like recent or intermediate memory, variable neighborhoods, or hashing scanning methods of the neighborhood. Then the method applied could be a Tabu Search, a Variable Neighborhood Search or any sophisticated hybrid heuristic search.

The metaheuristic strategy includes the decision on how to update the reference set taking into account the state of the evolution. The procedure must also realize when the reference set does not change and seek to diversify the search by generating a new set of individuals for the population. In addition, the metaheuristic includes the stopping criterion for the whole search procedure. Then the set of best individual used in the reference set is provided by the method; not only the best one. Usual stopping conditions are based on allowing a total maximum computational time or a maximum computational effort since the last improvement. The computational effort is measured by the number of iterations, number of local searches or real time.

The steps of the scatter search metaheuristic may be summarized as follows:

Scatter Search Metaheuristic

1. Create an initial population of individuals.
2. Generate a reference set from the population.
3. Select a subset from the reference set.
4. Apply a combination procedure to the subset.
5. Apply an improvement to the combinations.
6. Update the reference set accordingly.
7. Repeat 3 to 6 until a new reference set is needed.
8. Repeat 2 to 7 until a population is needed.
9. Repeat 1 to 8 until the stop criterion is met.

Figure 2: Scatter Search metaheuristic.

Therefore, the Scatter Search strategy is implemented by using 6 procedures and 3 stopping criteria to solve an optimization problem. These six procedures are based on 5 components as explained later.

The procedures are the following:

1. **The Initial Population Creation Procedure**. This procedure creates a random initial population of good and diverse individuals.

2. **The Reference Set Generation Procedure.** This procedure chooses the best *representative* individuals in the population to be included in the reference set.

3. **The Subset Selection Procedure.** This procedure gets *good* subsets of individuals in the reference set to apply the next combination procedure.

4. **The Individual Combination Procedure**. This procedure *intelligently* combines the individuals in each selected subset to get new individuals.

5. **The Improvement Individual Procedure.** This is an specialized procedure that improves the new individuals.

6. **The Reference Set Updating Procedure**. A procedure, provided with parameters used to modulate the intensification and/or diversification, updates the reference set by deciding when and how the obtained improved individuals are included in the reference set replacing some individuals already in it.

Note that the two procedures for constructing and updating the reference set are considered as a single component from the inspiration point of view that are related with the changes in the reference set. The usual *Reference Set* update procedure consists in applying the *Reference Set Generation* component to the union of the previous *Reference Set* and the individuals obtained by improving the combined solutions instead of the whole population.

In addition to these procedures, the metaheuristic involves three stopping procedures that implement the criteria to decide when to go to an above step.

1. **New Reference Set Criterion**. The first criterion decides when to generate a new reference set from the population.

2. **New Population Criterion**. The second criterion decides when to generate a new initial population.

3. **Termination Criterion**. Finally, the third criterion decides when to stop the whole search.

3. SCATTER SEARCH COMPONENTS

This section summarizes characteristics of the implementation of the components of the Scatter Search for the problems of optimally selecting a set of items from a universe. The fixed-sized selection problems include those optimization problems where the solutions are set of items with a given size. The solutions are sets represented as individuals by they characteristic function or other suitable data structure. Two of the most common problems are the *p*-median problem [16] and the feature selection problem [15]. The Scatter Search has been implemented and parallelized for those problems as can be seen in [3] and in [5], respectively (see also [4]).

The main characteristics of the five standard components of the Scatter Search are the following.

1. Initial Population Creation.

To implement this first Scatter Search component, one needs to choose first the size of the population based on the characteristics of the problems and data. So, for the feature subset selection problem, the individual space size depends on the number of features of the problem. Therefore, the size of the initial population is fixed depending on the number of features. We used in [5] $PopSize = d^2$, where d is the number of features. Several strategies can be applied for constructing random initial individuals. With a constructive procedure where the constructed individual depends on an initial point, it is applied from several starting points to get different individuals. The *Improvement Method* can be then applied to each individual obtained by the previous method. With the purpose of evaluating the dispersion among individuals, a distance between them is considered. This distance can be naturally defined by the size of the symmetric set difference ($Dist(X,Y) = |X - Y| + |Y - X|$) or using a nature-inspired distance between the items. Then, using a natural matching between the items of both individuals (since they have equal sized) the distance between them is the sum of the distances between paired items. Given a previously fixed size *PopSize* for the population *Pop*, α is the proportion of these *PopSize* individuals that are selected according to the objective function value. The remainder individuals up to *PopSize* can be obtained by an natural scoring procedure. For each individual X let the score be defined by

$$Score(\mathrm{X}) = Obj(X) - \beta\, Dist(X,Pop)$$

where $Dist(X,Pop)$ is the minimum distance between X and a individual already included in the population and β is a fixed factor. Then, the $(1-\alpha)PopSize$ best individuals according to $Score(.)$ are included in the population *Pop*. In order to build an individual, tacking into account the *diversification* objective, another vector of weights of the items to be included in the individuals can be used. These must weights indicate the quality of the possible items to be added to the individuals. For instance, let U be the set of items with the highest weights; i.e., a previously fixed number of items or those with score greater than a given threshold. Then, a possible *diversification* generation method consists of iteratively selecting at random one of the $|U|$ best possible items according their quality.

2. **Reference Set Generation**.

The generation of the reference set is usually done by selecting *RefSetSize*1 individuals from the best individuals and *RefSetSize*2 disperse individuals taking into account the above *RefSetSize*1 individuals (*RefSetSize* = *RefSetSize*1 + *RefSetSize*2) and some natural-inspired diversification measure. For instance, after including the best *RefSetSize*1 individuals in *RefSet*, one can iteratively include in the *RefSet* the most diverse individual with respect to the set of individuals already in *RefSet*, repeating this procedure *RefSetSize*2 times.

3. **Subset Selection.**

The selection of a subset for applying the combination usually consists in selecting all the subsets of fixed size r. For most of the implementation $r = 2$ is used. However, in order to avoid repetitions of combinations when some individuals of the reference set do not vary, information on the combinations performed is kept.

4. **Combination.**

Since there are usually only two individuals to be combined, the combination methods are very similar to the crossover operations. The only difference is that instead of doing random selections of the components from the parents, some heuristic rule is used to select the best (or one of the best) component from the parents. These methods are usually extended for more than two individuals in straightforward way. In order to combine good characteristics of the selected individuals to get the combined individuals several natural-inspired strategies can be applied. One general possible strategy consists in applying a constructive method like those explained above in the subpopulations resulting from the union of the individuals to be combined. A reduced version only considers those items that have appeared in good individuals. In order to apply this version some statistics on the items of the good individuals must be used.

5. **Improvement**.

The improvement component can be any procedure that performs a local search to a given individual. The most usual methods are those based on interchanges moves, which replace a item in the individual by a item out of the individual. These moves are the same kind of operations that mutations in Genetic Algorithms.

3. CONCLUSIONS

The Scatter Search is an evolutionary metaheuristic that is usually presented as a not natural inspired one, like Genetic Algorithm and others. However its components can be designed using natural inspiration. Moreover most of the implementations appeared in the literature have nature-inspired elements in its components or are very similar to those used in standard nature-inspired metaheuristics. The main difference is, unlike the strategies of operation rules usual in Genetic Algorithms; the search for a local optimum in a Scatter Search is guided, with less room for randomization.

REFERENCES

[1] Campos, V., Glover, F., Laguna, M., Martí, R., 2001, “An experimental evaluation of a scatter search for the linear ordering problem”, Journal of Global Optimization 21 pp. 397-414.

[2] Feo, T.A., and Resende, M.G.C., 1995, “Greedy randomized adaptive search procedures”. Journal of Global Optimization, 6, pp. 109-133.

[3] García-López, F.; Melián-Batista, B.; Moreno-Pérez, J.A.; Moreno-Vega, J.M., 2003, “Parallelization of the scatter search for the p-median problem”, Parallel Computing 29 pp. 575-589.

[4] García-López, F., García-Torres, M., Melián-Batista, B., Moreno-Pérez, J.A., Moreno-Vega, J.M., 2004 “Parallel Scatter Search”. In “Parallel Metaheuristics.” E. Alba (ed.), Wiley, Chapter 10

[5] García-López, F., García-Torres, M., Melián-Batista, B., Moreno-Pérez, J.A., Moreno-Vega, J.M., 2005 "Solving Feature Subset Selection Problem by a Parallel Scatter Search", European Journal of Operations Research (to appear).

[6] Glover, F., 1977 "Heuristics for Integer Programming using Surrogate Constraints", Decision Sciences 8, 156-166.

[7] Glover, F., 1994, "Tabu Search for Nonlinear and Parametric Optimization (with Links to Genetic Algorithms)", Discrete Applied Mathematics 49 pp. 231-155.

[8] Glover, F., 1998, "A Template for Scatter Search and Path Relinking", in "Lecture Notes in Computer Science",1363, J.K. Hao, E. Lutton, E. Ronald, M. Schoenauer, D. Snyers (Eds.), pp. 13-54.

[9] Glover, F., 1999, "Scatter Search and Path Relinking", in D. Corne, M. Dorigo, F. Glover (Eds.) "New Ideas in Optimisation", Wiley.

[10] Glover, F., Laguna, M., Martí, R., 2000, "Fundamentals of Scatter Search and Path Relinking" Control and Cybernetics, 39 pp. 653-684.

[11] Glover, F., M. Laguna, R. Martí,, 2002, "Scatter Search", in "Theory and Applications of Evolutionary Computation: Recent Trends", A. Ghosh, S. Tsutsui (Eds.) Springer-Verlag.

[12] Goldberg, D.E., 1989, "Genetic Algorithms in Search, Optimization and Machine Learning", Addison Wesley.

[13] Holland, J., 1975 "Adaptation in Natural and Artificial Systems", University of Michigan Press, Ann Arbor.

[14] Laguna, M.; Martí, R., 2003, "Scatter Search: Methodology and Implementation in C", Kluwer Academic Press.

[15] Mitchell, T. 1997, "Machine Learning", Series in Computer Science, McGraw-Hill.

[16] Maldenivic, N., Brimberg, J., Hansen, P., Moreno Pérez, J.A., 2005 "The p-median problem: A survey of metaheuristic approaches" European Journal of Operations Research, to appear.

[17] Resende, M., and Ribeiro, C., 2003, "Greedy Randomized Adaptive Search Procedures". In: Glover, Fred, and Kochenberger, Gary A. (eds.): "Handbook of Metaheuristics". (International Series in Operations Research & Management Science, vol. 57. Kluwer, chap. 8.