technische universität
dortmund

A New Clustering-Based Evolutionary Algorithm for Real-Valued Multimodal Optimization

Catalin Stoean, Mike Preuss,
Thomas Bartz-Beielstein,
Ruxandra Stoean

Algorithm Engineering Report
TR09-2-007
Sep. 2009
ISSN 1864-4503

Faculty of Computer Science
Algorithm Engineering (Ls11)
44221 Dortmund / Germany
http://ls11-www.cs.uni-dortmund.de/

A New Clustering-Based Evolutionary Algorithm for Real-Valued Multimodal Optimization

Catalin Stoean[1], Mike Preuss[2], Thomas Bartz-Beielstein[3], and Ruxandra Stoean[1]

[1] University of Craiova, A. I. Cuza, No. 13, 200585, Craiova, Romania
{catalin.stoean, ruxandra.stoean}@inf.ucv.ro
[2] Technische Universität Dortmund, Otto-Hahn, No. 14, 44227, Dortmund, Germany mike.preuss@tu-dortmund.de
[3] Cologne University of Applied Sciences, Steinmüllerallee, No. 1, 51643, Gummersbach, Germany bartz@gm.fh-koeln.de

Summary. Solving multimodal optimization tasks (problems with multiple global/local optimal solutions) by the state-of-the-art evolutionary algorithms (EAs) presumes separation of a population of individuals into subpopulations, each connected to a different optimum, with the aim of maintaining diversity for a longer period of time. Instead of using the typical separation that uses depends on a radius, present work proposes the employment of a clustering technique in order to distribute the candidate solutions to different species. Additionally, the proposed method corrects the separation by means of a mechanism that verifies the topological placement of the individuals in the fitness landscape with the purpose of connecting each species to a different optimum. The best individuals from each subpopulation are preserved from one generation to another in order to assure the conservation of the species. The method is applied on a set of benchmark functions that exhibit various properties, under multiple parameter settings, and the results demonstrate its great potential, especially of coping with relatively difficult problems under a limited budget of fitness evaluations.

1 Introduction

It is not unusual for an evolutionary heuristic to encounter difficulties when dealing with a multimodal problem: On the one hand, it might not detect the best solution because it got stuck into a local one and, on the other hand, it does not provide a set of the most *important* optima. The loss of diversity (genetic drift) represents the main issue when dealing with problems that exhibit a high number of optima. This is the reason why diversity has to be maintained for a longer period of time during evolution and that is achieved through the population separation into species.

Usually, the individuals that are closer than a given radius with respect to a specific metric are gathered within the same species (or subpopulation). In the best case scenario, each subpopulation is connected to a different peak in the problem landscape. However, that is impossible for a radius-based method, unless the attraction basins of the different peaks in the fitness landscape of the problem to be solved are equally sized and equidistant. It is hard to imagine that a real-world application would possess such properties. Moreover, a radius for separating the population also means that a user has to find a proper value for it, fact that is not straightforward and it greatly depends on the problem to be answered.

Within current work, a clustering technique is employed with the aim of gaining a good separation of the individuals into subpopulations. It uses distances and the fitness of the points to be clustered, but no radius parameter is involved. In order to correct the separation provided by the clustering method, an additional helping mechanism is used for determining whether the various species are connected to different hills in the landscape or not.

Next section presents a brief survey of other evolutionary methods for multimodal real-valued optimization problems, while the third outlines the proposed technique. Section four describes the experimental analysis on the application of the proposed technique for the optimization of a set of benchmark functions and the paper encloses with the concluding ideas.

2 Previous Approaches for Multimodal Optimization

Avoiding premature convergence (blocking into a local optimum) represents the key to developing a powerful EA for dealing with multimodal problems. This task is achieved by allowing diversity within the population for extended epochs and thus keeping contact with the promising regions within the landscape.

Many techniques have been proposed for improving performance when tackling multimodal problems. One of the most used (radius-based) evolutionary heuristics for multimodal optimization is the *fitness sharing* [5]. The preservation of diversity is conducted through the separation of the population according to niches. These are formed by joining individuals that lie closer than a given radius to one another. The main drawback of this method is represented by the existence of the threshold whose value very much relates to the specific problem. In this respect, Deb and Goldberg proposed a manner of computing the value for the radius that leads to the formation of subpopulations [2]. It is regularly embraced in most experimental designs for such EA methodologies. It uses the radius of the smallest hypersphere containing feasible space, which is given by (1).

$$r = \frac{1}{2}\sqrt{\sum_{i=1}^{D}(x_i^u - x_i^l)^2}. \tag{1}$$

In (1), D represents the number of dimensions of the problem at hand and x_i^u and x_i^l are the upper and the lower bounds of the i-th dimension. Knowing the number of optima N_G and being aware that each niche is enclosed by a D-dimensional hypersphere of radius r, the niche radius σ_{share} can be estimated as (2).

$$\sigma_{share} = \frac{r}{\sqrt[D]{N_G}}. \tag{2}$$

Note, however, that the employment of equation (2) for computing the value of the radius presumes that one must know in advance how many optima there exist within the fitness landscape.

Other methods like the island or diffusion models [4] that confine the way individuals interact have been imagined for the same purpose of keeping diversity more and better within the population. Although they do not have a radius threshold, these do not focus on the particularities of the landscape, but rather on the restrictions upon individuals when recombination is applied.

The crowding technique [3] was also developed as a method of maintaining diversity. New obtained individuals replace only similar ones in the population. However, there are some issues that have arisen around this methodology, as it suits only a limited set of multimodal problems and the value for the population size parameter very much depends on the considered task and the number of attraction basins that exist.

There have been previous attempts to use clustering in conjunction with EAs with the aim of targeting multimodal optimization. In [11] an adaptive clustering algorithm is introduced, meant to avoid the a priori estimation of the species separation radius. However, although the radius value does not appear anymore, it is necessary to define the minimum and the maximum radii of a cluster. A multi-population approach that also uses clustering for species separation is introduced in [8]. The main drawback is that the clustering technique brings two more parameters and one of them is a radius that separates the clusters.

The technique that is currently introduced partitions the population into species and it does not employ a radius parameter for that, instead, more efficiently, it connects each species to a different optimum by taking advantage of the fitness landscape and vicinity.

3 Nearest-Better Algorithm for Multimodal Optimization

Proposed nearest better separation evolutionary algorithm (NBSEA) has as its main engine a canonical EA that is endowed with a clustering technique. The latter has the goal of separating the population into species. Moreover, in order to maintain the contact with the promising regions from the search space, the conservation of the most prolific individual (which will be further on referred as *seed*) within each subpopulation is ensured. Additionally, a mechanism that verifies whether two candidate solutions lie within the same attraction basin or not is used in order to shape the subpopulations according to the fitness landscape of the problem to be solved. Both the clustering technique and the mechanism used for the verification of whether two solutions follow the same peak are heuristics, therefore they are combined in order to complement each other. In order to clarify each step, the following subsections will present every separate tool and finally the entire NBSEA is assembled in 3.3.

3.1 Nearest-Better Clustering

The clustering mechanism does not use any radius for separating the species, but it is based on the idea that the best individuals in different basins of attraction are more distant from each other as compared to the double of the mean distance of all individuals to their nearest better neighbors. The name of the technique is derived from this *nearest neighbor* principle it is based on. For each point, it searches for the closest neighbor, in terms of (Euclidean) distance, that is better in terms of fitness. The longest edges – those higher than $\phi \cdot mean$(lengths of all edges) – are removed and the prototypes for each cluster are represented by those individuals that do not connect to others. Clusters are thus formed around these (dominating) individuals. The approach has only one additional parameter to be tuned, with 2 being a good default value [10]. The nearest-better clustering (NBC) technique is thoroughly described in Algorithm 1.

The NBC algorithm is applied on the population of the EA after each generation in order to detect the dominating individual from each cluster (that corresponds further on to the species). The separation is not always optimal and, in order to detect whether there exist more than one cluster that lies within the same attraction basin, the seeds of the detected clusters are next verified two by two with the topological method that is further on described.

3.2 Detect Multimodal Mechanism

A mechanism to verify if two points lie within the same attraction basin or not was introduced within the multinational algorithm in [6]. It generates a number of equally distant interior points (this positive integer being the only parameter of the mechanism) between the two candidates it is applied for and, if there exists any interior point with the fitness smaller than the worse of the two initial ones, it is concluded that they follow different peaks. Otherwise, they are said to be situated on the same hill.

Algorithm 1 The Nearest-Better Clustering Algorithm

```
Require: A population of individuals $x$.
Ensure: Clusters of individuals: number, grouping, prototypes.
  for $i = 1$ to $pop\_size$ do
    compute distances from $x[i]$ to all individuals
  end for
  for $i = 1$ to $pop\_size$ do
    find nearest individual that is fitter than $x[i]$, i.e. $x[j]$;
    if found then
      $edge[i] = j$;
    else
      $edge[i] = 0$;
    end if
  end for
  $m = avg_{i=1}^{pop\_size}(distance(x[i], x[edge[i]]))$;
  for $i = 1$ to $pop\_size$ do
    if $edge[i] \neq 0$ and $distance(x[i], x[edge[i]]) > \phi \cdot m$ then
      $edge[i] = 0$;
    end if
  end for
  return the prototypes – $x[i]$, where $edge[i] = 0$ – and membership to clusters – $x[i] \in cluster[j]$, where $edge^n[i] = prototype[j]$, $i = 1, 2, ..., pop\_size$, $j = 1, 2, ..., noOfClusters$;
```

Algorithm 2 The Detect Multimodal mechanism for two individuals x and y

```
Require: Two individuals $x$ and $y$.
Ensure: $TRUE$ if $x$ and $y$ track different optima and $FALSE$ otherwise.
  $i = 1$;
  $found$ = FALSE;
  while $i <$ number of gradations and not $found$ do
    for $j = 1$ to number of dimensions do
      $interior_j = x_j + (y_j - x_j) \cdot gradation_i$;
    end for
    if $f(interior) < min(f(x), f(y))$ then
      $found$ = TRUE;
    end if
  end while
  return $found$;
```

The method is described in detail in Algorithm 2. The *number of gradations* coincides with the number of interior points and $gradation_i$, $i \in \{1, 2, ...,$ *number of gradations*$\}$, are equidistantly generated within $[0, 1]$. In a nutshell, the algorithm returns *true* if the two points follow different optima and $false$ if they lie within the same attraction basins.

3.3 Nearest-Better Separation Evolutionary Algorithm

The general outline of the currently proposed approach is presented in Algorithm 3. After the initialization of the population, the evolution progresses within a repetitive process. The NBC method (described in Algorithm 1) is applied in order to separate the population into species. Before moving forward, the $detect - multimodal$ mechanism is applied for the found cluster prototypes (that correspond to the species seeds). If there exist two (or more) seeds that follow the same peak, then their species are unified into one and the fittest of the seeds is kept as its representative. Early experimentation demonstrated the need to set an upper limit for the number of seeds as for test cases that are highly multimodal all the population risks to be blocked from the early generations into prototypes and therefore no exploration would be further achieved. In all our experimentation, the limit was set to 20% of the population size. Therefore, within Algorithm 1 the number of seeds, and species respectively, cannot be higher than the previously mentioned value.

Next, in the first generation, the selection operator is applied to the entire population and it is afterwards followed by the variation operators, recombination and mutation. If two individuals from the same species recombine, the offspring is considered to belong to the same group. In other recombination cases and in the situation when an individual is obtained via mutation, the offspring are considered $free$, meaning that they are not attached to any subpopulation.

As the selection and variation operators may omit and/or alter the master individuals of some species, a process of integration of the seeds that were previously detected occurs at this stage. Before inserting an individual, it is checked whether it already exists in the population, as it is introduced only in the case when there is no other copy of it. When such a seed is reinserted, it replaces the worst individual in its species or, in case its subpopulation was eliminated, it substitutes the worst member of the population that is not a master individual. The steps are described in Algorithm 4. The marking process takes place in order to avoid the deletion of previously inserted seeds.

The free individuals are then integrated to the existing species. The seeds are taken ascendingly according to the (Euclidean) distance to each free individual and it is tried, using the $detect - multimodal$ mechanism, to see if they follow the same peak or not. In the affirmative case, a free individual is assigned to the species, whose seed targets the same optimum as itself. When a free individual does not follow the same peak as any of the existing seeds, it is set as a new master individual and it creates its new own species.

Algorithm 3 Structure of the NBSEA

```
Require: A search/optimization problem
Ensure: The set of seeds
begin
Initialize population;
t = 0;
while stop condition is not met do
    Identify species seeds using Algorithm 1;
    Correct species through the use of Algorithm 2;
    if t > 0 then
        Reconsider old seeds;
    end if
    Apply mating selection;
    Apply recombination;
    Apply mutation;
    Integrate the seeds into resulting population;
    Assimilate the free individuals to the species;
    t = t + 1;
end while
return the set of seeds
end
```

Algorithm 4 Seeds conservation procedure within NBSEA

```
Require: The current population P
Ensure: The population that contains the seeds
begin
Mark all individuals in $P$ as unprocessed;
for every $s$ in $Seeds$ do
    if $s$ does not already exist in $P$ then
        Take worst unprocessed $w$ from P, such that $s_{ID} = w_{ID}$;
        if $w$ exists then
            if $f(w) < f(s)$ then
                $w = s$;
            end if
        else
            Take worst unprocessed $w$ in $P$;
            $w = s$;
        end if
        Mark $w$ as processed;
    end if
end for
return the population with the integrated seeds
end
```

Starting from the second generation, the previous findings are securely kept and the new gained information is updated. Right after seeds determination, the masters from the previous generation are checked against the newly found ones, once again in ascending order according to the distance, as the ones that are closer to each other are more likely to follow the same optimum. If there are seeds from the previous generation that follow different peaks ($detect - multimodal = true$), they are added to the current population and further set as seeds. They replace the worst individuals in the population that are not seeds.

As the $detect - multimodal$ mechanism is expensive with regard to the spent fitness evaluations, it is important to observe that it is always the NBC that is first applied within NBSEA and is followed by the verification method, as the latter is used only to correct the obtained separation.

The algorithm may stop after a fixed number of fitness evaluations is consumed (this is the case used in the experiments of the current paper), or after a fixed number of generations, or when the peaks are found with a desired accuracy.

4 Experimental Results

The set of test cases includes one unimodal function for testing whether the method can still cope with such simple tasks, two functions with many optima, all to be detected, and three other instances with one global optimum and many misleading local ones surrounding it. The aim of the tests is to observe how the proposed approach behaves under a very low budget of fitness evaluations and how the results improve when the investment rises. The results regard the peaks accuracies, the number of found peaks and the number of basins that are detected within the given landscape.

4.1 Test Functions

The simplest considered function is the Sphere ($F1$ in the enumeration below), it possesses only one optimum and is employed for 10 variables. Six-hump camel back ($F2$) is considered for 2 variables, and has 6 optima, equal two by two, that reside on a very smooth surface. Waves function is also considered for 2 variables and it has 10 optima (1 global) to be found, it is asymmetric and some peaks are difficult to find as they lie on the border or on flat hills. Keane's Bump Problem ($F4$), taken for 20 variables, is one case that has a strong relation to a real-world application from engineering design. It has a very bumpy surface and the global optimum is given by the product constraint. Shifted Rastrigin ($F5$) is taken for 2 and 10 variables, it has one global optimum and many local ones around. Finally, rotated hybrid composition function ($F6$) will be considered for the same number of dimensions as the latter, it is a very noisy case and it is obtained as the result of the combination of several functions: Its definition is too large to be included here and it can be found in [7]. The last two functions are part of the collection of test cases that were used as benchmarks for the Congress on Evolutionary Computation 2005 and more information about them, including encodings, can be found in [7]. The definitions for the described functions are itemized below.

$F1(x,y) = x^2 + y^2, -5.12 \leq x, y \leq 5.12.$

$F2(x,y) = -((4-2.1x^2+\frac{x^4}{3})x^2+xy+(-4+4y^2)y^2)$, $-1.9 \leq x \leq 1.9, -1.1 \leq y \leq 1.1.$

$F3(x,y) = (0.3x)^3 - (y^2 - 4.5y^2)xy - 4.7cos(3x - y^2(2+x))sin(2.5\Pi x))$, $-0.9 \leq x \leq 1.2, -1.2 \leq y \leq 1.2.$

$$F4(\overrightarrow{x}) = \frac{|\sum_{i=1}^{D} cos^4(x_i) - 2 * \prod_{i=1}^{D} cos^2(x_i)|}{\sqrt{\sum_{i=1}^{D} i * {x_i}^2}}, \ 0 \leq x_i \leq 10, \text{ subject to } \prod_{i=1}^{D} x_i > 0.75 \text{ and } \sum_{i=1}^{D} x_i < \frac{15 * D}{2}.$$

$$F5(\overrightarrow{x}) = \sum_{i=1}^{D}(x_i^2 - 10cos(2\Pi x_i) + 10) + f\_bias, \ -5 \leq x_i \leq 5.$$

$F6$ corresponds to function $F21$ in [7].

4.2 Task

Observe how NBSEA behaves for a number of fitness evaluations that is gradually raised from 1000 up to 30000, when applied to the functions that were introduced in the previous subsection.

4.3 Experimental Setup

An optimum is considered to be found if there exist at least one individual that has the difference in fitness evaluation to the actual peak closer than 10^{-1}. There are two measures considered for every function in the test suite and each is computed as the average value over 30 repeated runs:

1. *Peak ratio* gives the ratio between the number of found peaks and the actual number of existing peaks.
2. *Peak accuracy* refers to the summed difference in module between the fitness values of the desired optima and of the individuals that are closest to them (3).

$$\text{peak acc.} = \sum_{i=1}^{\#peaks} \mid f(peak_i) - f(x) \mid . \tag{3}$$

For each involved parameter, 30 Latin Hypercube Designs (LHDs) [1] are considered for generating equally distant values within reasonable intervals. Population size is taken between 10 and 200, mutation and recombination probabilities in [0, 1], the number of interior points between 1 and 10, while the mutation strength is dependant on the domains of definition, i.e. between [0, 5] for $F2$ and $F3$, [0, 15] for F4 and F5 with 2 variables and [0, 30] for the rest.

Concerning the used operators, tournament selection, intermediate recombination and mutation with normal perturbation are employed. The stop conditions are considered as 10^3, $2 \cdot 10^3$, $5 \cdot 10^3$, $7 \cdot 10^3$, 10^4 fitness evaluation calls and, from this point on, several steps obtained by adding 5000 evaluations up to 30000.

4.4 Results and Visualization

Obtained results for the first three functions in the best parameter configuration out of the 30 LHDs are illustrated in Figure 1.

$F1$ function is considered for 10 variables and the results prove that the smallest budget is not enough to conquer the existing peak. However, starting from 5000 evaluations, the peak is detected in all 30 repeats of the most prolific LHD.

The results appear to be substantially better for $F2$ when the number of fitness evaluations reaches 10^4 and beyond. The function possesses two local optima that have very low values as compared to the rest. They lie on narrow hills that are only slightly fitter than their surroundings and they are also positioned at the margins of the intervals. Therefore, it is very important that there are candidate solutions initialized in those regions in order to preserve (and improve) them. In this respect, a large value for the population size is necessary, even if that assumes the spending of a high number of fitness evaluations. Moreover, the number of interior points also has to be relatively high in order to detect the small differences in the fitness landscape.

For $F3$ the quality of the results progresses together with the increase in the fitness evaluation calls. However, it is worth mentioning that almost a half of the peaks are detected even for the lowest budget.

$F4$ is a very complex case, it is considered for 20 variables, and even the highest considered budget of fitness evaluations is not high enough to expect a very good performance. However, among the 30 LHDs, those that yield the best results are the ones that have the smallest values for the population size, in order to avoid spending evaluation calls and, at the same time, to prolong the evolutionary cycle. In this respect, it is the same configuration that gave

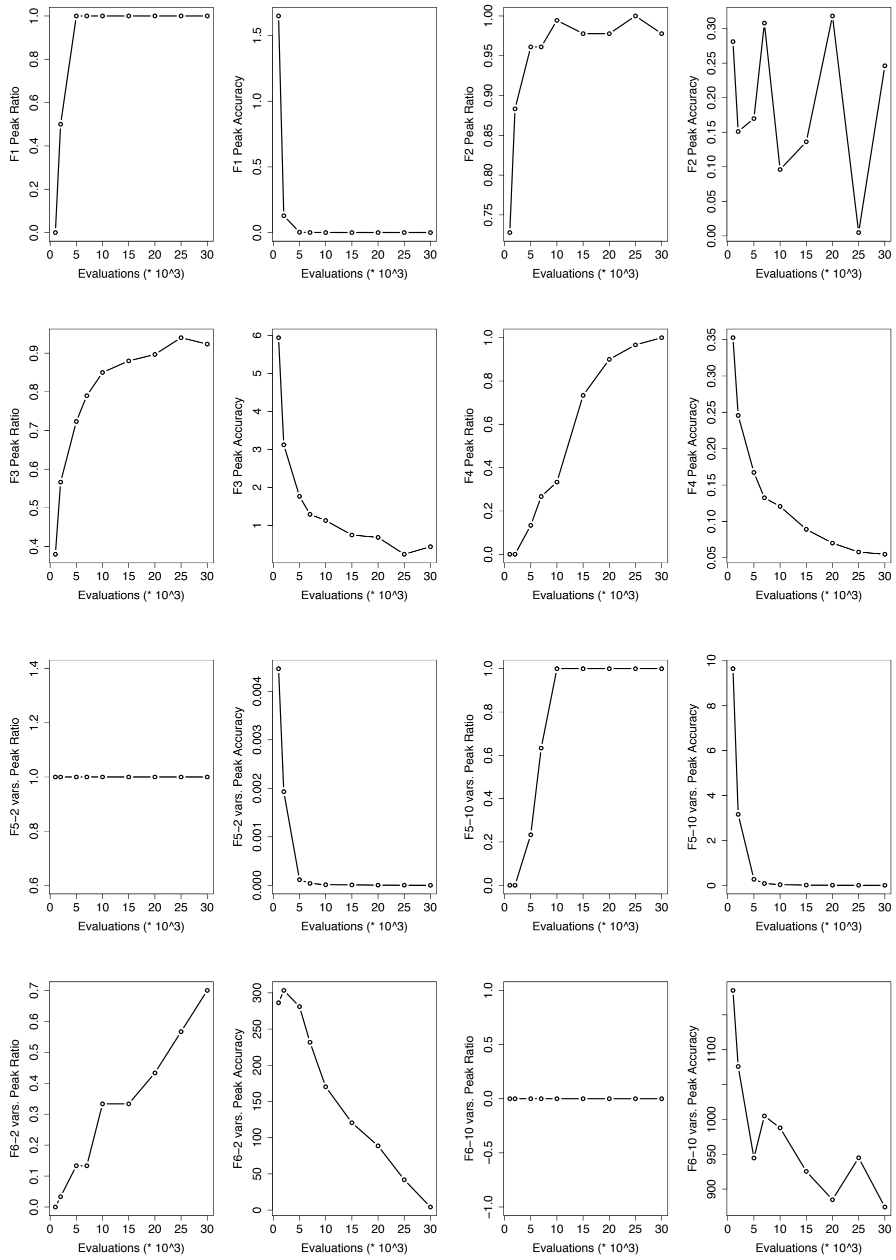

Fig. 1. Average results over 30 repeats of the best LHD configuration, obtained by NBSEA for the benchmark functions. For each case, first graphic contains the peak ratio (best value is 1), while the second illustrates the obtained peak accuracy (best value is 0). Each plotted results are obtained for 1000 up to 30000 fitness evaluation calls.

the best result for all stop conditions, one that has the population size 8, a high number of interior points, 15, and a high mutation strength of 29.1.

For $F5$ with 2 variables the global peak is found with the desired accuracy in all stop condition circumstances. For the same measurement, 10^4 evaluations were necessary for $F5$ with 10 variables for getting the best result. It is important to observe that relatively good results as concerns the peak accuracy are achieved starting with 5000 evaluations for $F5$ with both considered complexities. Afterwards, in both situations, there are only small improvements in peak accuracy (of the order 10^{-2}).

For $F6$ with 2 variables the results quality raises together with the evaluation investment. As a result of the CEC 2005 competition for function optimization, there no technique has been capable to detect the global optimum when $F6$ was considered for 10 variables. The proposed technique was not able to get too close to the global optimum, therefore the peak ratio graphic is not informative, but the progress of the NBSEA can be observed for the peak accuracy. As the function is very noisy, the method does not have a clear way towards improvement in dependence to the investments. However, there is some movement towards the peak when increasing the period of evolution.

The quality of the results could increase if more than 10^4 evaluation calls are considered. In average, the runtime of NBSEA takes 0.04 seconds for 10^3 fitness evaluation calls on a computer with an Intel Core 2 Extreme, 2 x 2.80 GHz processor and a RAM memory of 4 GB.

5 Conclusions and Future Directions

A new evolutionary technique for multimodal real-valued optimization that employs a clustering method for separating the population into species, nearest better separation evolutionary algorithm, is herein proposed.

The NBSEA gives very promising results, also for very economic investments regarding the number of fitness evaluation calls. However, it can be observed that, when more evaluations are used, the results improve accordingly. When this is allowed, then both the population size and the number of interior points can be taken higher; the highest the latter is, the more accurate the separation between subpopulation is.

It would be interesting to further investigate how the ϕ parameter within NBC influences the results and to compare them with other state-of-the-art evolutionary techniques, all under the same constraints. A parameter analysis would be also of great significance for the future work.

References

1. Bartz-Beielstein T. (2006) Experimental Research in Evolutionary Computation – The New Experimentalism, ser. Natural Computing Series, Berlin, Springer

2. Deb, K., Goldberg, D. E. (1989) An Investigation of Niche and Species Formation in Genetic Function Optimization, 3rd International Conference on Genetic Algorithms, Morgan Kaufman, San Mateo, California: 42–50
3. DeJong K. A. (1975) An analysis of the behavior of a class of genetic adaptive systems, Ph.D. dissertation, University of Michigan, Ann Arbor
4. Eiben A. E., Smith J. E. (2003) Introduction to Evolutionary Computing, Springer–Verlag, Berlin Heidelberg New York
5. Goldberg, D. E., Richardson, J. (1987) Genetic Algorithms with Sharing for Multimodal Function Optimization, 2nd International Conference on Genetic Algorithms, Lawrence Erlbaum Associates, Hillsdale, NJ: 41–49
6. Ursem, R., K. (1999) Multinational Evolutionary Algorithms, Proceedings of the Congress of Evolutionary Computation (CEC-99), vol. 3: 1633–1640
7. Suganthan P. N., Hansen N., Liang J. J., Deb K., Chen Y.-. P., Auger A., Tiwari S. (2005) Problem definition and evaluation criteria for the cec 2005 special session on real-parameter optimization, Technical Report, Nanyang Technological University, Singapore
8. Streichert F., Stein G., Ulmer H., Zell A. (2004) A Clustering Based Niching EA for Multimodal Search Spaces, Artificial Evolution, Springer Berlin / Heidelberg: 293–304
9. Bennett K. P., Blue J. (1997) A Support Vector Machine Approach to Decision Trees, R.P.I Math Report No. 97–100 Rensselaer Polytechnic Institute Troy, New York
10. Preuss, M., Schoenemann, L., Emmerich, M. (2005) Counteracting Genetic Drift and Disruptive Recombination in $(\mu\overset{+}{,}\alpha)$-EA on Multimodal Fitness Landscapes, 7th Annual Conference on Genetic and Evolutionary Computation, ACM Press: 865-872
11. Yin X., Germany N. (1993) A fast genetic algorithm with sharing using cluster analysis methods in multimodal function optimization, Proceedings of the International Conference on Articial Neural Nets and Genetic Algorithms: 450–457