A Clustering Based Niching Method for Evolutionary Algorithms

Felix Streichert[1], Gunnar Stein[2], Holger Ulmer[1], and Andreas Zell[1]

[1] Center for Bioinformatics Tübingen (ZBIT), University of Tübingen, Sand 1, 72074 Tübingen, Germany,
streiche@informatik.uni-tuebingen.de,
http://www-ra.informatik.uni-tuebingen.de

[2] Institute of Formal Methods in Computer Science (FMI), University of Stuttgart, Breitwiesenstr. 20/22, D-70565 Stuttgart, Germany,
http://www.informatik.uni-stuttgart.de/ifi/fk/index_e.html

1 Clustering Based Niching

We propose the Clustering Based Niching (CBN) method for Evolutionary Algorithms (EA) to identify multiple global and local optima in a multimodal search space. The basic idea is to apply the biological concept of species in separate ecological niches to EA to preserve diversity. To model species we use a multi-population approach, one population for each species. To identify species in a EA population we apply a clustering algorithm based on the most suitable individual geno-/phenotype representation. One of our goals is to make the niching method as independent of the underlying EA method as possible in such a way that it can be applied to multiple EA methods and that the impact of the niching method on the EA mechanism is as small as possible.

CBN starts with a single primordial unclustered population P_0. Then the CBN-EA generational cycle is entered. First for each population P_i one complete EA generation of evaluation, selection and reproduction is simulated. Now CBN starts with the differentiation of the populations by calling the clustering algorithm on each P_i. If multiple clusters are found in P_i, it is splits into multiple new populations. All individuals of P_i not included in the clusters found are moved to P_0 as straying loners. To prevent multiple populations to explore the same niche CBN uses representatives (e.g. a centroid) of all populations $\mathrm{P}_{i>0}$ to determine if populations are to be merged. To stabilize the results of the clustering algorithm we currently reduce the mutation step size within all clustered populations $\mathrm{P}_{i>0}$. A detailed description of the CBN model can be found in [2].

Of course the performance of CBN depends on the clustering algorithm used, since this algorithm specifies the number and kind of niches that can be distinguished. We decided to use the density-based clustering [1] which can identify an a priori unknown number of niches of arbitrary size, shape and spacing.

This multi-population approach of CBN replaces the global selection of a standard EA with localized niche based selection and mating. This ensures the survival of each identified niche if necessary. Also each converged population $\mathrm{P}_{i>0}$ directly designates a local/global optimum.

Table 1. Mean of found optima, in parentheses the number of evaluations needed.

	M0		M1		M2		M3	
No. of optima	**5**		**5**		**6**		**10**	
MS-HC	4.80	(6.000)	4.90	(6.000)	4.52	(6.000)	8.70	(6.000)
Sharing	4.66	(6.000)	4.54	(6.000)	1.98	(6.000)	8.40	(6.000)
MN-GA(W)	4.83	(355.300)	5.00	(355.300)	5.60	(812.300)	8.98	(1.221.600)
MN-GA(N)	4.94	(355.300)	4.99	(355.300)	3.91	(812.300)	9.80	(1.221.600)
CBN-ES	5.00	(6.000)	4.64	(6.000)	3.94	(6.000)	8.10	(6.000)

2 Results and Conclusions

We examined a CBN Evolution Strategy (ES), a standard ES with fitness sharing with an additional hill-climbing post-processing step and a μ-multi-start hill-climber (MS-HC). We used a $(\mu + 2 \cdot \mu)$-ES, $\mu = 100$ and $T = 60$ generations as default settings. We compared these algorithms the Multinational GA (MN-GA) on four real-valued two-dimensional test functions [3]. The performance is measured by the number of optima each algorithm has found, averaged over fifty runs. An optimum o_j is considered as found if $\exists\ x_i \in P_{t=T} \mid \|x_i, o_j\| \leq \epsilon = 0.005$, with the final population $P_{t=T} = \bigcap_i P_{i,t=T}$ in the case of CBN.
Tab. 1 shows that the MN-GA needs much more fitness evaluation than the ES based methods. It shows also that the MS-HC performs well on these simple test functions, so does Sharing in combination with the HC post-processing. Although the parameters for MS-HC and Sharing where optimized for each problem, the CBN-ES proves to be competitive with default parameters.
The advantages of CBN are that is does not alter the search space, that it is able to find niches of arbitrary size, shape and spacing and that it inherits all properties of the applied EA method, since it does not significantly interfere with the EA procedure. There are a number of extensions that can further enhance the CBN. First applying a population size balancing in the case of unevenly sized areas of attraction. Second using a greedy strategy of convergence state management to save function evaluations if a population $P_{i>0}$ is converged.

References

1. M. Ester, H.-P. Kriegel, J. Sander, and X. Xiaowei. A density-based algorithm for discovering clusters in large spatial databases with noise. In E. Simoudis, J. Han, and U. Fayyad, editors, *2nd International Conference on Knowledge Discovery and Data Mining*, pages 226–231, Portland, Oregon, 1996. AAAI Press.
2. G. Stein. Verteiltes dynamisches Nischenmodell fuer JavaEvA. (in German), Diploma Thesis at the Institute of Formal Methods in Computer Science (FMI), University of Stuttgart, Germany, 2002.
3. R. K. Ursem. Multinational evolutionary algorithms. In P. J. Angeline, Z. Michalewicz, M. Schoenauer, X. Yao, and A. Zalzala, editors, *Proceedings of the Congress on Evolutionary Computation*, volume 3, pages 1633–1640, Mayflower Hotel, Washington D.C., USA, 6-9 1999. IEEE Press.