A Novel Clustering Fitness Sharing Genetic Algorithm

Xinjie Yu

State Key Lab of Power Systems, Dept. of Electrical Engineering,
Tsinghua University, Beijing 100084, China
yuxj@tsinghua.edu.cn

Abstract. The hybrid multimodal optimization algorithm that combines a novel clustering method and fitness sharing method is presented in this paper. The only parameter required by the novel clustering method is the peak number. The clustering criteria include minimizing the square sum of the inner-group distance, maximizing the square sum of the inter-group distance, and the fitness value of the individuals. After each individual has been classified to the certain cluster, fitness sharing genetic algorithm is used to find multiple peaks simultaneously. The empirical study of the benchmark problems shows that the proposed method has satisfactory performance.

1 Introduction

Many real world optimization problems are multimodal in essence. Thus it is very convenient for product designers and decision makers to select one solution from several candidates. Though Simple Genetic Algorithm (SGA) has been proved very useful in optimization, machine learning and many other industrial application areas, it can only converge to a single peak in the search space [1, 2]. Theoretical analyses and numerical experiments show that with finite population size and weak selective pressure, the single convergence, named as genetic drift, may occur [3].

Fitness sharing genetic algorithm is one of the most useful methods for multimodal optimization problems, which reduces the individual's fitness value according to the distance between individuals [4]. Improvements on standard fitness sharing genetic algorithm include clearing procedure suggested by Petrowski [5], dynamic niche sharing suggested by Miller and Shaw [6], and adaptive niching via coevolutionary sharing suggested by Goldberg and Wang [7]. These algorithms need the peaks radii before performing fitness sharing method, which is hard to estimate for some problems.

The organization of this paper goes as follows. Section 2 surveys the clustering methods combined with fitness sharing genetic algorithms and presents the new clustering method. Section 3 studies the suggested approach by optimizing the benchmark multimodal problems. Conclusions appear in the last section.

2 The Novel Clustering Fitness Sharing Genetic Algorithm

2.1 Surveys on Clustering Methods for Fitness Sharing Genetic Algorithm

Both standard fitness sharing genetic algorithm and its improvements need to identify the centers and the peaks radii. Clustering methods have been used to grouping the data

L. Wang, K. Chen, and Y.S. Ong (Eds.): ICNC 2005, LNCS 3611, pp. 1072 – 1079, 2005.

for a long time. Combining clustering methods with fitness sharing genetic algorithm may improve the adaptability and the speed of the algorithm.

Yin and Germay presented a fast genetic algorithm with sharing scheme using adaptive MacQueen's KMEAN clustering algorithm [8]. The adaptive KMEAN starts generating k initial centers of the cluster according to the fitness value of the individuals. Then the centers merge and receive new individuals using $d_{\min}$ and $d_{\max}$.

In Lin, Liu and Yang's paper, a new cluster technique is proposed for automatic and adaptive identification of the locations and the sizes of the clusters in genetic algorithms with fitness sharing [9]. Although no cluster number and cluster radii are required, the algorithm needs several predefined parameters.

Torn's clustering method has been used by many researchers [10]. The method takes the best individual as the first cluster center. Hanagandi and Nikolaou use Torn's clustering method to do global optimization [11]. After several generations of GAs, Torn's clustering method is carried out and cluster centers are found.

2.2 The New Clustering Method

Every clustering method has its own cluster criteria. The cluster criteria adopted by the new clustering method contain the square sum of the inner-group distance and square sum of the inter-group distance [12].

Given N individuals and k clusters, the square sum of inner-group distance can be defined as follows:

$$J_1 = \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left\| x_i - m_j \right\|^2 \tag{1}$$

where n_j is the number of individuals of the cluster j, x_i is the i'th individual in the cluster j, and m_j is the center of the cluster j. The smaller the inner-group distance criterion is, the better the clustering result is.

The square sum of the inter-group distance criterion can be defined as follows:

$$J_2 = \sum_{j=1}^{k} \left\| m_j - m \right\|^2 \tag{2}$$

where m is the center of the total individuals. The larger the inter-group distance criterion is, the better the clustering result is.

In this paper, only the multimodal optimization problems with equal fitness values of the peaks are concerned. The suggested clustering method combines minimizing the square sum of the inner-group distance, maximizing the square sum of the inter-group distance, and the fitness value information of individuals in an easy heuristic way. The novel clustering method need the number of peaks k beforehand and can be illustrated in Fig. 1.

The first part considers the fitness value information of individuals. The better the individual is, the earlier it will be treated. The second part tries to maximize the square sum of the inter-group distance and the third part tries to minimize the square sum of the inner-group distance. The last part combines the novel clustering method with the fitness sharing method.

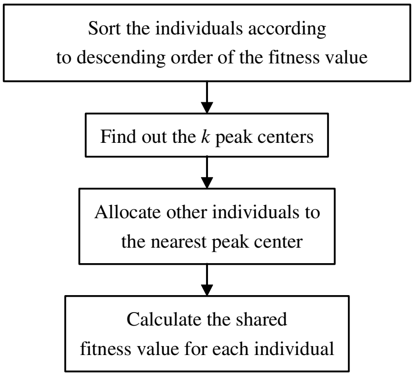

Fig. 1. The novel clustering method

The specific algorithm for finding the k peak centers can be expressed as follows:

1. The first individual is the first *confirmed* peak center, and mark it as *selected.*
2. Select k individuals which are not marked as selected orderly from the population.
3. Calculate the distances from the k individuals to the confirmed peak centers.
4. For each selected individual, appoint its *representative distance* as its nearest distance to the confirmed peak centers.
5. Select the largest representative distance from the k representative distances and point the corresponding individual as the next confirmed peak center, and mark it as selected
6. If the number of the confirmed peak centers is less than k, go to step 2.

The specific algorithm for allocating other individuals can be expressed as follows:

1. For each individual, which is not marked as selected, calculate its distances to the k confirmed peak centers.
2. Select the shortest distance and allocate the individual to the corresponding confirmed peak center, market the individual as *allocated*, and record which confirmed peak center it belongs to.
3. Go to step 1 until all the individuals are allocated.

In the novel clustering method, a niche is a peak. The specific algorithm for calculating the shared fitness value for each individual can be expressed as follows

1. For each niche, find out the number of individuals (m_i) in it. (Note that every individual can only be allocated to one niche.)
2. For each individual in that niche, its niche count is m_i, and its shared fitness value is f / m_i (f is its raw fitness value)
3. Go to step 1 until every individual gets its shared fitness value.

The time complexity of distance calculation of this method can be considered as follows. The number of distance calculation in finding the peak centers is $k+2k+\ldots+(k-1)\,k$.

The number of distance calculation in allocating individuals is (N-1)k. Generally speaking, N>>k, so the total time complexity of distance calculation is O(Nk), which is the same as that of the adaptive KMEAN clustering method[8] and much less than O(N^2) of the standard fitness sharing method [13]. Additional analyses show that the suggested method does not need steps to merge clusters and identify the cluster centers, so the total calculation time of the new method may be smaller than that of the adaptive KMEAN clustering method.

The predetermined parameter k is easy to obtain in some circumstance, especially when the problem is to calculate the root of algebraic equation. If there is not any clue about k, users can increase k continuously.

2.3 The Novel Clustering Fitness Sharing Genetic Algorithm

The novel clustering fitness sharing genetic algorithm can be illustrated in Fig. 2.

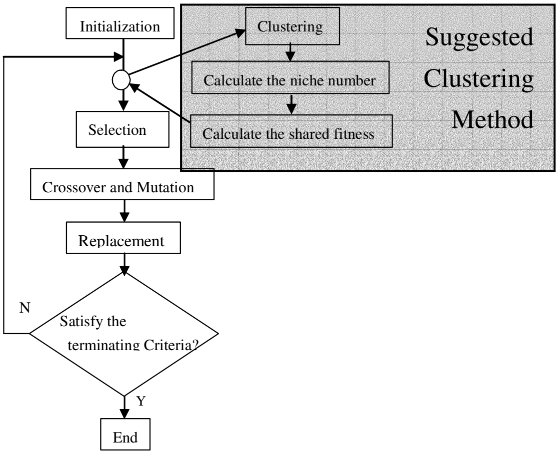

Fig. 2. The novel clustering fitness sharing genetic algorithm

3 Empirical Studies on the Suggested Algorithm

The algorithm is compared with other multimodal genetic algorithms using 3 benchmark problems. The algorithms include the standard fitness sharing (SH), the fitness sharing with adaptive KMEAN clustering method (KMEAN), the new clustering fitness sharing method (NEW).

3.1 Benchmark Problem Specification

Only maximization problems with same peak height are considered as the test problems. The sequence of problem represents the increase of multimodality.

Problem I can be expressed as follows [14]:

$$F1(x) = \sin^6\left[5\pi\left(x^{0.75} - 0.05\right)\right] \tag{3}$$

The domain of the problem is [0,1]. There are 5 unequally spaced peaks with same height. The maxima are located at x values of 0.080, 0.247, 0.451, 0.681, and 0.934. All peaks are of height 1.0.

Problem II can be expressed as follows [15]:

$$F2(x) = 2500 - \left(x(1)^2 + x(2) - 11\right)^2 - \left(x(1) + x(2)^2 - 7\right)^2 \tag{4}$$

The domain of the problem is [-6,6]*[-6,6]. There are 4 unequally spaced peaks with same height. The maxima are located at x values of (3.5844, -1.8481), (3,2), (-2.8051,3.1313), and (-3.7793, -3.2832). All peaks are of height 2500.0.

Problem III is the massive deceptive problem and can be expressed as follows [16]:

$$f\left(x_0,\cdots,x_{29}\right) = \sum_{i=0}^{4} u\left(\sum_{j=0}^{5} x_{6i+j}\right) \tag{5}$$

where $\forall\ x_k \in \{0,1\}, k = 0,\cdots,29$. $u(s)$ is defined as follows:

$$u(s) = \begin{cases} 1 & s \in \{0,6\} \\ 0 & s \in \{1,5\} \\ 0.360384 & s \in \{2,4\} \\ 0.640576 & s = 3 \end{cases} \tag{6}$$

The search space of the problem is of size 10^9, the number of peaks is of size 10^6, but the number of global peaks is only 32. Operations like crossover and mutation are likely to generate local peaks. The maxima of problem III are located at x values of (000000, 000000, 000000, 000000, 000000), …,(111111, 111111, 111111, 111111, 111111). All peaks are of height 5.

These benchmark problems are the most common ones while testing the multimodal genetic algorithms. As can be seen from Section 2, the proposed clustering based fitness sharing genetic algorithm has no difficulty while scaling with increasing dimensions.

3.2 Parameters and Performance Criteria

To compare the three algorithms fairly, the parameters of each algorithm are set to be equal, as shown in Table 1. Mating restriction strategy is adopted in SH [13]. Twenty runs are carried out with different initial population generated at random and the average of these runs is taken for comparison.

Table 1. The parameters for the algorithms to solve 3 benchmark problems

Problem	I	II	III
Selection Type	SUS [17]	SUS [17]	SUS [17]
Crossover Type	Single Point	Single Point	Single Point
Crossover Probability	0.9	0.9	0.9
Mutation Probability	0.05	0.05	0.05
Scaling Type	No	Power Law Scaling[18]	Power Law Scaling[18]
Distance Type	Euclidean Distance	Euclidean Distance	Hamming Distance
Population Size	60	100	400
Maximum Generation	50	50	100
Chromosome Length	30	15	30
Sigma (SH)	0.1 [13]	4.2426 [13]	6 [16]
$d_{\min}$, $d_{\max}$ (KMEAN)	0.05, 0.1 [8]	1.5, 3	2, 6
Initial Cluster Number (KMEAN)	5	4	32
Peak Number (NEW)	5	4	32
Run Time	20	20	20
Convergence Criterion	$h<0.02$	$h<0.832$	At Peaks

where h is the distance between the individual and the peak nearest to it.

Four performance criteria are used to evaluate the algorithms. The number of global peaks maintained by algorithms is criterion I. It is very clear that the larger the criterion I is, the better the algorithm is. The chi-square-like number is the criterion II [13]. Smaller chi-square-like number means more uniform distribution of the population. The number of individuals resided in the global peaks is the criterion III. The large number of criterion I and the large number of criterion III represent good solution. But the large number of the criterion III and the small number of criterion I represent premature convergence. The run time is the criterion IV, measured by second. The calculation is carried out with MATLAB on the computer with 466MHz Celeron CPU and 128MB memories. The less the run time is, the better the algorithm is.

3.3 Results

Table 2 lists the average results of 20 runs using 3 algorithms on 3 benchmark problems respectively.

Problem I represents the multimodal problem with different peaks radii, and it is quite easy to solve. So all three algorithms' results are nearly the same except the run time. SH’s run time is almost two times of that of NEW, and KMEAN’s run time is in intermediate level.

Problem II represents the multimodal problem with flat fitness landscape. KMEAN seems to be inconvenient for this kind of problem, and NEW is efficient and excellent. Although SH can find all the peaks, its run time is three times of that of NEW.

Table 2. The computation result of the algorithms

	Problem I			Problem II			Problem III		
Criteria	SH	KMEAN	NEW	SH	KMEAN	NEW	SH	KMEAN	NEW
I	4.8	4.6	4.8	4.0	3.6	3.9	31.9	2.0	30.9
II	3.7	3.9	3.0	5.8	8.3	3.0	12.1	79	7.0
III	45	44	52	53	33	82	260	400	345
IV	71.1	55.5	37.5	188.3	115.8	61.2	6046.2	3597.0	3941.7

Problem III represents the multimodal problem with massive deception, and is very hard to solve. SH achieves the best solution result but the run time cost is rather expensive. KMEAN can only find two global peaks because its search ability is weaker than the other two. NEW can maintain 30.9 global peaks with lower chi-squire-like criterion and less run time.

It is very clear that the overall champion is NEW. SH is good for its solution quality, but suffers from long run time. KMEAN is quicker than SH, but its search ability is weak.

4 Conclusions

In this paper, a novel clustering method based on the criteria of minimizing the inner-group distance and maximizing the inter-group distance has been presented. The clustering method also considers the fitness information of individuals and is very easy to combine with the standard fitness sharing genetic algorithm. Properties of the suggested clustering method are discussed as follows:

- The time complexity of distance calculation of the new clustering method is $O(Nk)$, which is the same as that of the adaptive KMEAN clustering method and much less than $O(N^2)$ of the standard fitness sharing method. The total calculation time of the new method is smaller than that of the adaptive KMEAN clustering method because it does not need to identify the cluster center and merge the clusters.
- The predetermined parameter of the new clustering method is the number of peaks, which is very convenient in some circumstances, especially in finding roots of algebraic equation.
- The solution quality of the novel clustering fitness sharing genetic algorithm is satisfactory for various kinds of multimodal benchmark problems.

References

1. Holland J.H. ︰ Adaptation in Natural and Artificial Systems. Ann Arbor, MI: University of Michigan Press (1975)
2. Goldberg, D.E.: Genetic Algorithms in Search, Optimization, and Machine Learning. New York: Addison-Wesley (1989)
3. Mahfoud, S.W.: Genetic Drift in Sharing Methods," In: Proceedings of the First IEEE Conference on Evolutionary Computation. Piscataway, NJ: IEEE Press (1994) 67–72
4. Goldberg, D.E. and Richardson J.: Genetic Algorithms with Sharing for Multimodal Function Optimization. In: Grefenstette, J.J. (eds.): Proceedings Of the Second International Conference on Genetic Algorithms and Their Applications. Hillsdale, NJ: Lawrence Erlbaum (1987) 41–49
5. Petrowski, A.: A Clearing Procedure as a Niching Method for Genetic Algorithms. In: Proceedings of the third IEEE Conference on Evolutionary Computation. Piscataway, NJ: IEEE Press (1996) 798–803
6. Miller, B.L. and Shaw, M. J.: Genetic Algorithms with Dynamic Niche Sharing for Multimodal Function Optimization. In: Proceedings of the third IEEE Conference on Evolutionary Computation. Piscataway, NJ: IEEE Press (1996) 786–791
7. Goldberg, D.E. and Wang, L.: Adaptive Niching via Coevolutionary Sharing. IlliGAL Report No. 97007. (1997)
8. Yin, X. and Germay, N.: A Fast Genetic Algorithm with Sharing Scheme Using Cluster Analysis Methods in Multimodal Function Optimization. In: Albrecht, R.F. (eds.): Proceedings of International Conference on Artificial Neural Nets and Genetic Algorithms. New York, Springer-Verlag (1993) 450–457
9. Lin C.Y, Liu J.Y, and Yang Y.J.: Hybrid Multimodal Optimization with Clustering Genetic Strategies. Engineering Optimization (1998) 30: 263–280
10. Torn A.: A Search-clustering Approach to Global Optimization. In: Dixon, S. (eds.): Towards Global Optimization, Vol. 2. North-Holland (1978) 49–70
11. Hanagandi V. and Nikolaou M.: A Hybrid Approach to Global Optimization Using A Clustering Algorithm in A Genetic Search Framework. Computers Chem. Engng. (1998) 22(12): 1913–1925
12. Everitt B.S. Cluster Analysis. 3rd ed. New York. John Wiley & Sons (1993)
13. Deb, K. and Goldberg, D.E.: An Investigation of Niche and Species Formation in Genetic Function Optimization. In: Schaffer, J.D. (eds.): Proceedings of the Third International Conference on Genetic Algorithms and their Applications. San Mateo, CA: Morgan Kaufmann (1989) 42–50
14. Sareni, B. and Krahenbuhl, L.: Fitness Sharing and Niching Methods Revisited. IEEE Transactions on Evolutionary Computation. Vol. 2, no. 3, (1998) 97–106
15. Harik G.: Finding multimodal solutions using restricted tournament selection. In: Eshelman, L. (eds.): Proceedings of the 6th. International Conference on Genetic Algorithms. San Mateo, CA: Morgan Kaufmann (1995) 24–31
16. Goldberg D.E., Deb K., and Horn J.: Massive Multimodality, Deception, and Genetic Algorithms. In: Manner R. and Manderick B. (eds.): Proceedings of the Second Conference on Parallel Problem Solving from Nature. Amsterdam: North-Holland (1992) 15–25
17. Baker, J. E.: Reducing Bias and Inefficiency in the Selection Algorithm. In: Grefenstette, J.J. (eds.): Proceedings Of the Second International Conference on Genetic Algorithms and Their Applications. Hillsdale, NJ: Lawrence Erlbaum (1987) 14–21
18. Yu, X. and Wang, Z.: The Fitness Sharing Genetic Algorithms with Adaptive Power Law Scaling. System Engineering Theory and Practice. 2002. Vol. 22, no. 2, 42–28