

Multimodal Estimation of Distribution Algorithm Based on Cooperative Clustering Strategy

Shanshan Huang¹, Houming Jiang²

1. Nari Group Corporation (State Grid Electric Power Research Institute), Nanjing, 210003
E-mail: huangshanshan@sgepri.sgcc.com.cn

2. Nari Group Corporation (State Grid Electric Power Research Institute), Nanjing, 210003
E-mail: jianghouming@sgepri.sgcc.com.cn

Abstract: Multimodal optimization is becoming more and more important. Among all the optimization algorithms, estimation of distribution algorithm (EDA) attracts more attention because of its simple principle and special mechanism. Unfortunately, it loses feasibility and effectiveness in multimodal problems because of premature convergence and the missing of multimodality-specific mechanism. So niching is adopted to help EDA to locate diverse promising solution regions in multimodal optimization. However, many of the niching techniques are either sensitive to parameters or cost a number of fitness evaluations. In this paper a new proposed fast cooperative clustering strategy is adopted to offer improved assistance for locating multiple optima based on both decision and target space information. Taking the advantage of EDA in preserving high diversity, this paper proposes a cooperative clustering based multimodal EDA (CMEDA). Integrated with new proposed cooperative clustering strategy, multimodal problems are divided into certain promising regions automatically. Then each cluster independently runs a separate optimizer in parallel to search promising regions carefully, which can avoid premature convergence greatly. Based on the results of clustering, the value of key parameter is determined by statistic information but not artificial setting. Then a balance between exploration and exploitation is achieved. Experimental results indicate that the proposed technique is an effective and efficient algorithm which can not only explores and exploits the promising regions in the search space effectively but also obtain the global optima superior to the typical multimodal EDA (MEDA).

Key Words: Estimation of Distribution Algorithm (EDA), Multimodal Optimization, Multiple Global Optima, Niching, Cooperative Clustering Strategy

1 INTRODUCTION

Multimodal optimization, which seeks multiple optima simultaneously, has attracted more and more attention in recent years [1]. Real-world problems like data mining [2], flowshop schedule [3], holographic design [4], and electromagnetic design [5] require to search more than one optima. However, different from finding just one optimum in single optimization [6], locating local optima simultaneously or multiple global is qualitatively more challenging [1].

For these problems, classical and typical evolutionary algorithms (EAs), such as differential evolution (DE) [7], genetic algorithm (GA) [8], and particle swarm optimization (PSO) [9], lose feasibility and effectiveness, because their overall learning and updating makes the population tend to converge toward one dominating candidate [9]. Recently, a new family of EAs—the estimation of distribution algorithm (EDA) [10], has emerged, which gains more attention because of its special mechanism and performance. It generates offspring by sampling from the probability distribution estimated from promising individuals [10].

Unfortunately, EDA loses feasibility and effectiveness in multimodal problems because of premature convergence and the missing of multimodality-specific mechanism [11]. Therefore, to locate multiple optima simultaneously using classical EAs, a multimodality-specific mechanism is necessary. Under this circumstance, niching is adopted to help EDA maintain a diverse population in multimodal optimization, which achieves this by partitioning the whole population into several subpopulations using techniques. However, many of the niching techniques such as clustering, crowding and speciation are either sensitive to parameters or cost a number of fitness evaluations [1]. For instance, [2] shows that k-means clustering is susceptible to cluster number K and [11] describes how crowding is sensitive to the crowding size.

In this paper, a new fast clustering strategy, Cooperative Clustering [12], is adopted to offer improved assistance for locating multiple optima based on both decision space and target space information. This strategy not only locates diverse promising solution regions, but also reduces the sensitivity of the used niching methods to the cluster size [12]. In order to take advantage of the new strategy, multimodal estimation of distribution algorithm based on cooperative clustering strategy (CMEDA) is proposed. In this algorithm multimodal problems are divided into certain promising regions automatically and then each cluster independently runs a separate optimizer individually. By

This work is supported by the company's self-financing technology projects under Grant 5246DR180002.

this means a more accurate multimodal model is built and global optimal solutions are gained so that the balance between exploration and exploitation is achieved. The simulation results based on CEC2013 niching methods for multimodal function optimization [13] validate that CMEDA outperforms the previously existing similar algorithm.

2 EDA and Cooperative Clustering Strategy

So far, multimodal algorithms are all based on DE [7], GA [8], or PSO [9], which exhibit various deficiencies, such as inefficiency in dealing with complex problems, where masses of local optima exist. Few attempts have been made to the development of EDA for multimodal application [10], despite EDA features substantial exploration with high diversity at the population level [10], which would be beneficial for multimodal optimization. In terms of such fact, this paper mainly focuses on EDA.

2.1 EDA

EDA is a special class of model-based evolutionary algorithms (EAs) that is characterized by the way of generating new solutions [6]. It produces new solutions for current generation according to a certain probability distribution which is generally estimated from the relatively superior solutions in previous generation, hoping that the structures of the promising solution regions can be captured and consequently employed to guide the search process [10]. EDA executes in the following iterative way:

Table 1. Main Procedures of EDA

Procedure	Operation
Selection	Select certain promising solutions
Modelling	Build a probabilistic model
Sampling	Sample certain new solutions on model
Replacement	Replacement solutions to form new population

EDA generally chooses truncation selection to select promising solutions, in which individuals are sorted according to their objective function values [6]. Then Gaussian model is used as the basic probability distribution model [6]. The n-dimensional joint probability distribution is factorized as a product of n univariate and independent normal distributions [10]. There are two parameters for each variable required to be estimated in each generation: the mean μ_i^k , and the standard deviation σ_i^k . The estimation of them are given as follows:

$$\hat{\mu}_i^k = \bar{x}_i^k = \frac{1}{M} \sum_{t=1}^M x_{t,i}^k \quad (1)$$

$$\hat{\sigma}_i^k = \sqrt{\frac{1}{M} \sum_{t=1}^M (x_{t,i}^k - \bar{x}_i^k)^2} \quad (2)$$

where $(x_{1,i}^k, x_{2,i}^k, \dots, x_{M,i}^k)$ are values of the i^{th} variable of the selected M parent solutions in the k^{th} generation.

EDA builds a Gaussian probability distribution model based on promising solutions of each generation, which has simple principle and special mechanism [6]. With the evolutionary,

the model can infinitely close to the actual distribution of single model problems [6]. Many researchers has claimed that it has outstanding performances on single model problems [10] because the model can infinitely close to the actual distribution of single model problems. Despite possessing analytical properties, the performance of Gaussian EDA is still undesirable while referring to multimodal problems [1]. EDA loses feasibility and effectiveness in multimodal problems because of premature convergence and the missing of multimodality-specific mechanism.

2.2 Cooperative Clustering Strategy

Here clustering algorithm is adopted to capture different promising solution regions of high-quality solutions, so we hope it has neither too many parameters nor too much calculation. However, these niching techniques either cost a number of fitness evaluations or require a large memory or introduce less sensitive parameters to partition the population into niches [14]. Then in this paper a new fast clustering strategy is proposed which is inspired by the thought of [15]. The clustering strategy is based on distance and density, which is mainly used in image processing area. Considering the characteristic of objective problems, some improvements of the mentioned strategy has been done to form the cooperative clustering strategy. It chooses solutions with better fitness and farther relative distance from other solutions as cluster centers, then parts the rest solutions into different clusters, which solves the parameter problem of existing clustering algorithms and the set of SD value.

(1) Computing the relative distance from all the solutions with better fitness

The basic distance is defined as (3).

$$\delta_{ij} = \sqrt{\|X_i - X_j\|^2 / D} \quad (3)$$

where δ_{ij} denotes the distance between i^{th} and j^{th} solution, X_i and X_j respectively represent the position of i^{th} and j^{th} solution, $\|\cdot\|$ denotes the distance and D is the dimension.

For the i^{th} weed, its distance matrix Δ_i contains the distance between itself and specific weeds whose fitness is better than the i^{th} weed. This can be described as (4). We predefined $\Delta_i = 0$.

$$\Delta_i = \{\delta_{ij} \mid f_i > f_j, i \neq j\} \quad (4)$$

After calculating the distance matrix of each weed, the minimum distance δ_i of each weed is defined as (5).

$$\begin{cases} \delta_i = \min(\Delta_i) \\ \delta_i = \max(\delta_i) \end{cases} (i > 1) \quad (5)$$

(2) Computing the distance threshold

The threshold of distance is shown in (6). Among the weeds with minimum distance, those who satisfy expression (7) are chosen as cluster centers. $weed(i)$ denotes the i^{th} weed.

$$\begin{cases} \delta_{\max} = \max(\delta_i) \\ \delta_{\min} = \min(\delta_i) \\ \delta_{\text{threshold}} = (\delta_{\max} - \delta_{\min}) \times 80\% \end{cases} \quad (6)$$

(3) Cluster centers and cluster members

$$\text{centers} = \{\text{weed}(i) | \delta_i > \delta_{\text{threshold}}\} \quad (7)$$

After determining the cluster centers, the rest solutions belong to the cluster whose center has the nearest distance.

(4) Plot the relative distance as a function of fitness

To illustrate the performance of this strategy, some experiments were done and the result is shown in Figure 1. It shows the result of choosing cluster centers of unimodal function 3 with 10 dimensions. As described above, cluster centers are those points who locate in the upper area. The rest solutions are then distribute to relevant cluster. The result indicates the effectiveness of DS-TS clustering strategy.

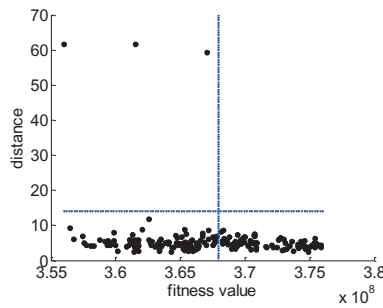


Figure 1. Result of choosing cluster centers

This self-learning mechanism classifies solutions within the same local region into a cluster so that one-sidedness of cluster decision is avoid. The calculation process does not require iteration so that it has low computational complexity [12]. The number of cluster is gained during the cluster process adaptively but not given in advance so that it reduces the sensitivity of cluster size. In conclusion, cooperative clustering strategy has comparative advantage over other existing niching algorithms.

3 Multimodal EDA Based on Cooperative Clustering Strategy

Without modification, an EDA is unsuitable for multimodal optimization because a number of global maxima are distributed across various areas surrounded by multiple local maxima [1]. In addition, some global optima are far away from one another, while some are very close to one another [16]. In such an environment, an EDA is more likely to converge toward one or a small number of global optima. So locating multiple global optima is the main objective of multimodal optimization [1], and hence is also the concern of this paper. As shown in the following Figure 2, the subregions in black cycles are promising areas to find max-optimal solutions, which locate in different regions. To cope with multimodal optimization efficiently, a dynamic cluster sizing strategy is absorbed in the niching methods for EDA. It is used to help EDA maintain a diverse

population in multimodal optimization by partitioning the whole population into subpopulations and each subpopulation is responsible for one area to locate one or a small number of optima [16].

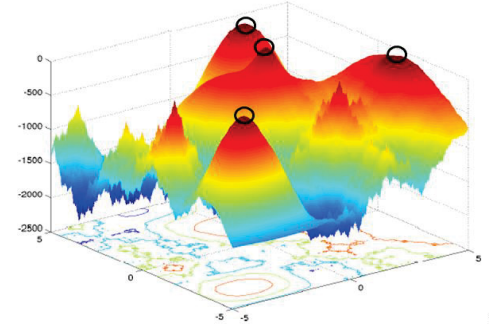


Figure 2. Multimodal optimization

3.1 Key Improvements of CMEDA

Based on the theory states above, the key points of CMEDA are listed as following:

1) Double space cooperative clustering strategy: This extracts features of problems adaptively based on both decision space and target space information and then decision space is divided into different subspaces.

2) The build of probability distribution model of each subspace: Different Gaussian models are built within each cluster. A more specific operation is that weeds in different clusters and different dimensions share different SD values which are determined by statistical information of better weeds in their own clusters but not artificial setting.

3) Model adjustment: Within each cluster, the model is adjusted to a more promising state which is closer to the statistical true model. The model center is shifted along a vector which points to the better one between the present best solution and that of the last generation and start at the other one. The offset is determined by a certain value which is directly collected to the SD value. On usual, the better the statistical center is, the smaller the certain value.

4) Sampling to reproduce new population: Sample new solutions based on distribution model. Within each subspace, individuals are picked out to constitute high-quality solutions according to cluster scale, which means the larger a cluster the more individuals will be picked.

5) Select best optima according to distance-fitness criterions: Based on distance-fitness criterions solutions with better fitness value and far apart from each other are choose as best global optima.

3.2 the Framework of CMEDA

To build a more appropriate model for multimodal problems, this paper adopted a cooperative clustering strategy which can capture different promising solution regions with the obtained solutions adaptively.

Taking the advantage of EDA in preserving high diversity, multimodal problems are divided into subspace automatically. After classifying, each species is evolved individually using the EDA optimizer cooperatively. EDA selects better solutions from initial population with truncation selection according to their objective function values. Then it builds a Gaussian probability distribution model based on promising solutions of each generation, which has simple principle and special mechanism. New solutions are sampled according to the constructed probabilistic model to update the population.

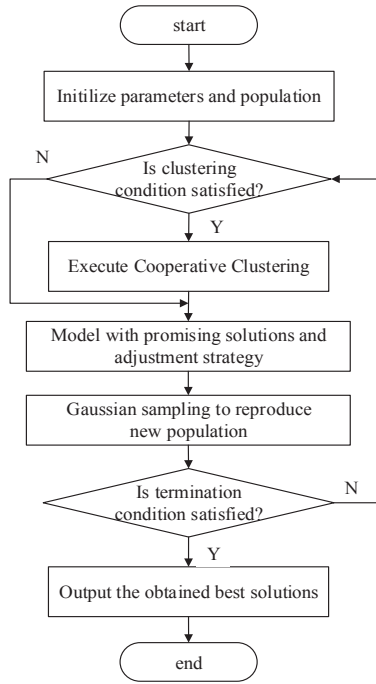


Figure 3. The framework of CMEDA

During this frame, multimodal problems are divided into several single model problems by clustering strategy and then each sub-problem is optimized with EDA cooperatively. The model built by this means is more accurate and comprehensive.

Based on the results of clustering, the value of key parameter is determined by statistic information but not artificial setting. Then a balance between exploration and exploitation is achieved so that computing resources is distributed to avoid premature convergence. At last, optima are selected based on a series of distance-fitness criterions to evaluate the performance.

4 Experiment studies

4.1 Benchmark Functions and Evaluation Protocols

To test the effectiveness of CMEDA, we conduct experiments on a widely used benchmark function set—the CEC'2013 multimodal function set [13] containing 20

functions, which are designed for the 2013 IEEE CEC special session on niching methods for multimodal optimization. To evaluate the performance of CMEDA and make fair comparisons with the state-of-the-art multimodal algorithms, peak ratio (PR) and success rate (SR) are selected as the evaluation protocols [13], which are also adopted in the corresponding competition on niching methods for multimodal optimization at the special session [1]. Given a fixed maximum number of function evaluations and a specific accuracy level ε , PR is defined as the percentage of the number of the global optima found out of the total number of global optima averaged over multiple runs [17]. SR measures the ratio of successful runs out of all runs and a successful run is defined as a run where all known global optima are found [13]. In the experiments [13], five different accuracy levels, namely: $\varepsilon = 1.0\text{E-}01$, $\varepsilon = 1.0\text{E-}02$, $\varepsilon = 1.0\text{E-}03$, $\varepsilon = 1.0\text{E-}04$, and $\varepsilon = 1.0\text{E-}05$, are adopted.

To make fair comparisons, the maximum number of fitness evaluations (Max_Fes) and the population size (NP) are set to the same for all compared multimodal methods as shown in Table 2.

Table 2. Parameter Settings [1]

Function	Max_Fes	NP
F1-F5	5.0E+04	80
F6	2.0E+05	100
F7	2.0E+05	300
F8-F9	4.0E+05	300
F10	2.0E+05	100
F11-F13	2.0E+05	200
F14-F20	4.0E+05	200

4.2 Benchmark Functions and Evaluation Protocols

In this part, CMEDA is compared with DE-Niching to test its effectiveness. The results are showed in Table 3 and are directly compared with a typical DE algorithm, which incorporates spatial information about the neighborhood of each potential solution to produce a niching formation [13]. It produces niching formation by incorporating the crowding technique to maintain a better population diversity and therefore to prevent premature convergence to an optimum [13]. Table 3 shows the comparison results between CMEDA and DE-Niching [1] with respect to PR and SR on all 20 functions at all accuracy levels and the best PR results are emphasized in bold. From the table, we can observe the following:

The two algorithms achieve the same performance on 6 functions (F1, F2, F3, F4, F5, F13) at all five accuracy levels. Besides, on F1–F5, they can locate all known global optima successfully at any accuracy level. On F13, both of them can find two thirds global optima at all accuracy levels.

Compared with DE-Niching, overall, CMEDA is slightly better on F6, F8, F11, F14, F16, F17, F18, and shows its great superiority to DE-Niching on F10 with finding all global optima.

Then, CMEDA can successfully find all global optima at all accuracy levels on F7, while DE-Niching can only locates less than a half global optima. On the last 5 functions, CMEDA performs better than DE-Niching on certain accuracy level while a bit poor on the others accuracy levels. In a conclusion, CMEDA can locate all global optima on five accuracy levels on 35% of 20 test functions and 75%

better than DE-Niching. In summary, we can conclude that from the perspective of locating more global optima as concerned in this paper, CMEDA is superior to DE-Niching so that it is promising for solving multimodal optimization problems.

Table 3. Comparison Results in PR and SR Between CMEDA and DE-Niching

Accuracy Level	F1				F2				F3				F4				F5			
	CMEDA		DE-Niching		CMEDA		DE-Niching		CMEDA		DE-Niching		CMEDA		DE-Niching		CMEDA		DE-Niching	
	PR	SR	PR	SR	PR	SR	PR	SR	PR	SR	PR	SR	PR	SR	PR	SR	PR	SR	PR	SR
1.0E-01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.0E-02	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.0E-03	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.0E-04	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.0E-05	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Accuracy Level	F6				F7				F8				F9				F10			
	CMEDA		DE-Niching		CMEDA		DE-Niching		CMEDA		DE-Niching		CMEDA		DE-Niching		CMEDA		DE-Niching	
	PR	SR	PR	SR	PR	SR	PR	SR	PR	SR	PR	SR	PR	SR	PR	SR	PR	SR	PR	SR
1.0E-01	0.50	0.00	0.45	0.00	1.00	1.00	0.35	0.00	0.14	0.00	0.11	0.00	0.08	0.00	0.10	0.00	1.00	1.00	1.00	1.00
1.0E-02	0.50	0.00	0.43	0.00	1.00	1.00	0.35	0.00	0.14	0.00	0.11	0.00	0.07	0.00	0.10	0.00	1.00	1.00	1.00	1.00
1.0E-03	0.50	0.00	0.44	0.00	0.83	0.00	0.35	0.00	0.14	0.00	0.11	0.00	0.07	0.00	0.10	0.00	1.00	1.00	0.99	0.98
1.0E-04	0.50	0.00	0.43	0.00	0.33	0.00	0.34	0.00	0.14	0.00	0.11	0.00	0.07	0.00	0.10	0.00	1.00	1.00	1.00	1.00
1.0E-05	0.50	0.00	0.00	0.00	0.06	0.00	0.33	0.00	0.14	0.00	0.11	0.00	0.07	0.00	0.09	0.00	1.00	1.00	1.00	1.00
Accuracy Level	F11				F12				F13				F14				F15			
	CMEDA		DE-Niching		CMEDA		DE-Niching		CMEDA		DE-Niching		CMEDA		DE-Niching		CMEDA		DE-Niching	
	PR	SR	PR	SR	PR	SR	PR	SR	PR	SR	PR	SR	PR	SR	PR	SR	PR	SR	PR	SR
1.0E-01	1.00	1.00	0.68	0.00	0.75	0.00	0.86	0.24	0.68	0.00	0.67	0.00	1.00	1.00	0.67	0.00	0.50	0.00	0.52	0.00
1.0E-02	0.83	0.00	0.67	0.00	0.75	0.00	0.84	0.22	0.67	0.00	0.67	0.00	0.67	0.00	0.67	0.00	0.50	0.00	0.54	0.00
1.0E-03	0.83	0.00	0.68	0.00	0.75	0.00	0.82	0.14	0.67	0.00	0.67	0.00	0.67	0.00	0.67	0.00	0.50	0.00	0.51	0.00
1.0E-04	0.83	0.00	0.67	0.00	0.75	0.00	0.82	0.16	0.67	0.00	0.67	0.00	0.67	0.00	0.67	0.00	0.50	0.00	0.50	0.00
1.0E-05	0.83	0.00	0.67	0.00	0.75	0.00	0.78	0.10	0.67	0.00	0.67	0.00	0.67	0.00	0.67	0.00	0.50	0.00	0.51	0.00
Accuracy Level	F16				F17				F18				F19				F20			
	CMEDA		DE-Niching		CMEDA		DE-Niching		CMEDA		DE-Niching		CMEDA		DE-Niching		CMEDA		DE-Niching	
	PR	SR	PR	SR	PR	SR	PR	SR	PR	SR	PR	SR	PR	SR	PR	SR	PR	SR	PR	SR
1.0E-01	0.67	0.00	0.67	0.00	0.38	0.00	0.35	0.00	0.43	0.00	0.40	0.00	0.25	0.00	0.23	0.00	0.13	0.00	0.13	0.00
1.0E-02	0.67	0.00	0.66	0.00	0.38	0.00	0.33	0.00	0.33	0.00	0.34	0.00	0.13	0.00	0.17	0.00	0.13	0.00	0.13	0.00
1.0E-03	0.67	0.00	0.66	0.00	0.38	0.00	0.30	0.00	0.33	0.00	0.32	0.00	0.13	0.00	0.15	0.00	0.13	0.00	0.13	0.00
1.0E-04	0.67	0.00	0.66	0.00	0.38	0.00	0.29	0.00	0.33	0.00	0.28	0.00	0.13	0.00	0.13	0.00	0.13	0.00	0.13	0.00
1.0E-05	0.67	0.00	0.66	0.00	0.38	0.00	0.29	0.00	0.33	0.00	0.25	0.00	0.13	0.00	0.13	0.00	0.13	0.00	0.12	0.00

5 Experiment studies

This paper has developed CMEDA to locate multiple global optima for multimodal optimization problems. EDA and niching are effectively utilized to realize the proposed algorithms [18]. Specially, the clustering-based niching tactics which based on decision space and target space are

used to enhance performance on multimodal problems. Their superior performance on multimodal problems is brought about by the cooperative clustering strategy and the model adjustment techniques developed in this paper.

The niching method of CMEDA is improved by developing a dynamic cluster-sizing strategy to afford a potential balance between exploration and exploitation, whereby

relieving CMEDA from the sensitivity to the cluster size. Differing from classical EDA to estimate the probability distribution of the whole population, CMEDA focuses on the estimation of distribution at the niche level and all individuals in each niche participate in the estimation of distribution of that niche. Further, the alternative usage of statistical information guides the adjustment of sample model within each cluster. Finally, the solution accuracy is enhanced through a new local search scheme based on Gaussian distribution with probabilities self-adaptively determined according to fitness values of solutions. The comparison results with respect to PR and SR between CMEDA and the state-of-the-art multimodal algorithm demonstrates the superiority and efficiency of CMEDA developed in this paper.

REFERENCES

- [1] Q. Yang, W. N. Chen, Y. Li, C. L. Philip Chen, X. M. Xu and J. Zhang, Multimodal Estimation of Distribution Algorithms, *IEEE Trans. Cybern.*, Vol.47, No.3, 636-650, 2017.
- [2] Y. Li, Using Niche Genetic Algorithm to Find Fuzzy Rules[J], *Proc. Int. Symp. Web Inf. Syst. Appl.*, Nanchang, China, 64-67, 2009.
- [3] Z. W. Sun and X. S. Gu, A novel hybrid estimation of distribution algorithm for solving hybrid flowshop scheduling problem with unrelated parallel machine[J], *Journal of Central South University*, Vol. 24, No. 8, 1779-1788, 2017.
- [4] Q. Ling, G. Wu and Q. Wang, Restricted evolution based multimodal function optimization in holographic grating design, *Proc. IEEE Congr. Evol. Comput.*, 789-794, Edinburgh, U.K., 2005.
- [5] D. Woo, J. Choi, M. Ali and H.-K. Jung, A novel multimodal optimization algorithm applied to electromagnetic optimization[J], *IEEE Trans. Magn.*, Vol. 47, No. 6, 1667-1673, Jun. 2011.
- [6] A. Zhou, J. Sun and Q. Zhang, An estimation of distribution algorithm with cheap and expensive local search methods, *IEEE Trans. Evol. Comput.*, Vol. 19, No. 6, 807-822, Dec. 2015.
- [7] Y. L. Li, Z. H. Zhan, Y. J. Gong, W. N. Chen and J. Zhang, Differential Evolution with an Evolution Path: A DEEP Evolutionary Algorithm[J], *IEEE Trans. Cybern.*, Vol. 45, No. 9, 1798-1810, Sep. 2015.
- [8] X. Zhang, J. Zhang, Y. Gong, Z. H. Zhang and W. N. Chen, Kuhn-Munkres Parallel Genetic Algorithm for the Set Cover Problem and Its Application to Large-Scale Wireless Sensor Networks[J], *IEEE Trans. Evol. Comput.*, Vol. 20, No. 5, 695-710, 2016.
- [9] Y. Gong, J. Li, Y. Zhou, Y. Li and S. H. Chung, Genetic learning particle swarm optimization, *IEEE Trans. Cybern.*, Vol. 46, No. 10, 2277, 2016.
- [10] P. Yang, K. Tang and X. Lu, Improving estimation of distribution algorithm on multimodal problems by detecting promising areas, *IEEE Trans. Cybern.*, Vol. 45, No. 8, 1438-1449, Aug. 2015.
- [11] R. Thomsen, Multimodal optimization using crowding-based differential evolution[C], in *Proc. IEEE Congr. Evol. Comput.*, Vol. 2, 1382-1389, USA, 2004.
- [12] S. S. Huang, Z. G. Ren, C. L. Sun and M. J. Wang, A Two Stages Invasive Weed Optimization via a New Clustering Strategy[C], *Genetic and Evolutionary Computation Conference Companion*, ACM, 1001-1005, 2016.
- [13] X. Li, A. Engelbrecht and M. G. Epitropakis, "Benchmark functions for CEC'2013 special session and competition on niching methods for multimodal function optimization," *Evol. Comput. Mach. Learn. Group, RMIT Univ.*, Melbourne, VIC, Australia, Tech. Rep., 2013.
- [14] W. Dong and X. Yao, NichingEDA: Utilizing the diversity inside a population of EDAs for continuous optimization, *Proc. IEEE Congr. Evol. Comput.*, 1260-1267, Hong Kong, 2008.
- [15] S. Das and P. N. Suganthan, Differential evolution: A survey of the state-of-the-art, *IEEE Trans. Evol. Comput.*, Vol. 15, No. 1, 4-31, Feb. 2011.
- [16] G. Tan, K. Bao and R. M. Rimiru, A composite particle swarm algorithm for global optimization of multimodal functions[J], *Journal of Central South University*, Vol. 21, No. 5, 1871-1880, 2014.
- [17] R. Thomsen, Multimodal optimization using crowding-based differential evolution[C], *IEEE Congress on Evolutionary Computation*, Vol. 2, 1382-1389, 2004.
- [18] W. Dong, T. Chen, P. Tino and X. Yao, Scaling up estimation of distribution algorithms for continuous optimization[J], *IEEE Transactions on Evolutionary Computation*, Vol. 17, No. 6, 797-822, 2013.