Applied Soft Computing 41 (2016) 479–487

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Not guaranteeing convergence of differential evolution on a class of multimodal functions

Zhongbo Hu [a,*], Qinghua Su [a], Xianshan Yang [a], Zenggang Xiong [b]

[a] School of Information and Mathematics, Yangtze University, Jingzhou, Hubei 434023, PR China
[b] School of Computer and Information Science, Hubei Engineering University, Xiaogan, Hubei 432000, PR China

ARTICLE INFO

Article history:
Received 4 January 2015
Received in revised form 2 December 2015
Accepted 3 January 2016
Available online 21 January 2016

Keywords:
Continuous optimization
Convergence in probability
Differential evolution algorithm
Drift analysis

ABSTRACT

The theoretical studies of differential evolution algorithm (DE) have gradually attracted the attention of more and more researchers. According to recent researches, the classical DE cannot guarantee global convergence in probability except for some special functions. Along this perspective, a problem aroused is that on which functions DE cannot guarantee global convergence. This paper firstly addresses that DE variants are difficult on solving a class of multimodal functions (such as the Shifted Rotated Ackley's function) identified by two characteristics. One is that the global optimum of the function is near a boundary of the search space. The other is that the function has a larger deceptive optima set in the search space. By simplifying the class of multimodal functions, this paper then constructs a Linear Deceptive function. Finally, this paper develops a random drift model of the classical DE algorithm to prove that the algorithm cannot guarantee global convergence on the class of functions identified by the two above characteristics.

1. Introduction

The differential evolution algorithm (DE) proposed by Storn and Price in 1995 [1] is a population-based stochastic parallel evolutionary algorithm. DE emerged as a very competitive form of evolutionary computing [2–4] and has got many practical applications, such as function optimization, multi-objective optimization, classification, scheduling and so on.

Since that theoretical studies benefit understanding the algorithmic search behaviors and developing more efficient algorithms, more and more researchers pay attention to the theoretical studies on DE with the popularity in applications. In 2005, Zielinski et al. [5] investigated in theory the runtime complexity of DE for various stopping criteria including a fixed number of generations (G_{max}) and maximum distance criterion (*MaxDist*). From 2001 to 2010, Zaharie [6–11], Dasgupta et al. [12,13] and Wang et al. [14] analyzed the dynamical behavior of DE's population from different perspectives, i.e., the statistics characteristics, the gradient-descent type search characteristics and stochastic evolving characteristics respectively. Recently some convergent DE algorithms [15–19] have developed.

This paper focuses on the convergence analyses of the classical DE. Several important conclusions on the convergence have been drawn. In 2005, Xu et al. [20] performed a mathematical modeling and convergence analysis of continuous multi-objective differential evolution (MODE) under certain simplified assumptions, and this work was extended in [21]. In 2012, Ghosh and Das et al. [22] used Lyapunov stability theorem to establish the asymptotic convergence behavior of a classical DE (DE/rand/1/bin) algorithm on a class of special functions identified by the following two properties, 1) the function has the second-order continual derivative in the search space, and 2) it possesses a unique global optimum in the range of search. In 2013, Hu et al. [23] proposed and proved a sufficient condition for global convergence of the modified DE algorithms. In 2014, Hu et al. [24] developed a Markov chain model of the classical DE and proved then that it cannot guarantee global convergence in probability. In a word, the classical DE cannot guarantee global convergence in probability except for some special functions.

Along this perspective, this paper does two works as follows:

- Firstly, this paper addresses that DE variants are difficult to solve a class of multimodal functions. By abstracting the characteristics of the class of functions, this paper then constructs a Linear Deceptive function which can simplify the theoretical analyses on DE.

* Corresponding author.
E-mail addresses: huzbdd@yangtzeu.edu.cn, huzbdd@126.com (Z. Hu).

http://dx.doi.org/10.1016/j.asoc.2016.01.001
1568-4946/

- This paper develops a random drift model of the classical DE algorithm to prove the conclusion that the algorithm cannot guarantee global convergence on a class of functions represented by the Linear Deceptive function.

The rest is organized as follows. Section 2 introduces the classical DE algorithm. As the research background of this paper, Section 3 presents a problem that many DE variants are difficult to solve a class of multimodal functions. Section 4 qualitatively analyzes the reason resulting the problem by using distribution characteristics of the trial population, and offers the proof idea of the main conclusion in the following sections. Sections 5 and 6 prove the main conclusion that the classical DE cannot guarantee global convergence on a class of multimodal functions by constructing a Linear Deceptive function and developing a random drift model of the classical DE. Finally the concluding remarks are presented in Section 7.

2. Classical differential evolution

DE is used for dealing with the continuous optimization problem. We suppose in this paper that the objective function to be minimized is $f(\vec{x}),\quad \vec{x}=(x_1,\ldots,x_n)\in\Re^n$, and the feasible solution space is $\Psi=\prod_{j=1}^{j=n}[L_j,U_j]$. The classical DE [1,3,26] works through a simple cycle of operators including mutation, crossover and selection operator after initialization. The classical DE procedures are described in detail as follows.

2.1. Initialization

The first step of DE is the initialization of a population of m n-dimensional potential solutions (*individuals*) over the optimization search space. We shall symbolize each individual by $\vec{x}_i^g=(x_{i,1}^g,x_{i,2}^g,\ldots,x_{i,n}^g)$, for $i=1,\ldots,m$, where $g=0,1,\ldots,g_{max}$ is the current generation and g_{max} is the maximum number of generations. For the first generation ($g=0$), the population should be sufficiently scaled to cover the optimization search space as much as possible. Initialization is implemented by using a random number distribution to generate the potential individuals in the optimization search space. We can initialize the jth dimension of the ith individual according to

$$x_{i,j}^0=L_j+rand(0,1)\cdot(U_j-L_j)$$

where $rand(0,1)$ is a uniformly distributed random number confined in the [0,1] range.

2.2. Mutation operators

After initialization, DE creates a donor vector $\vec{v}_i^g$ corresponding to each individual $\vec{x}_i^g$ in the gth generation through the mutation operator. Several most frequently referred mutation strategies are presented as follows:

DE/rand/1:

$$\vec{v}_i^g=\vec{x}_{r_1}^g+F(\vec{x}_{r_2}^g-\vec{x}_{r_3}^g);$$

DE/best/1:

$$\vec{v}_i^g=\vec{x}_{best}^g+F(\vec{x}_{r_1}^g-\vec{x}_{r_2}^g);$$

DE/current-to-best/1:

$$\vec{v}_i^g=\vec{x}_i^g+F(\vec{x}_{best}^g-\vec{x}_i^g)+F(\vec{x}_{r_1}^g-\vec{x}_{r_2}^g);$$

DE/best/2:

$$\vec{v}_i^g=\vec{x}_{best}^g+F(\vec{x}_{r_1}^g-\vec{x}_{r_2}^g)+F(\vec{x}_{r_3}^g-\vec{x}_{r_4}^g);$$

DE/rand/2:

$$\vec{v}_i^g=\vec{x}_{r_1}^g+F(\vec{x}_{r_2}^g-\vec{x}_{r_3}^g)+F(\vec{x}_{r_4}^g-\vec{x}_{r_5}^g);$$

where $\vec{x}_{best}^g$ denotes the best individual of the current generation, the indices $r_1,r_2,r_3,r_4,r_5\in S_r=\{1,2,\ldots,m\}\setminus\{\,i\,\}$ are uniformly random integers mutually different and distinct from the running index i, and $F\in(0,1]$ is a real parameter, called *mutation* or *scaling factor*.

If the jth element of $\vec{v}_i$ is infeasible (i.e. out of the boundary), it is reset as the following rule, called Symmetrical Mode Rule.

$$v_{i,j}=\begin{cases}2L_j-v_{i,j} & \text{if}\quad v_{i,j}<L_j\\ 2U_j-v_{i,j} & \text{if}\quad v_{i,j}>U_j\end{cases}$$

2.3. Crossover operator

Following *mutation*, the *crossover* operator is applied to further increase the diversity of the population. In crossover, the target vectors, $\vec{x}_i^g$, are combined with elements from the donor vector, $\vec{v}_i^g$, to produce the trial vector, $\vec{u}_i^g$, using the *binomial crossover*,

$$u_{i,j}^g=\begin{cases}v_{i,j}^g & \text{if}\quad rand(0,1)\le CR\quad or\quad j=j_{rand}\\ x_{i,j}^g & \text{otherwise}\end{cases}$$

where $CR\in(0,1)$ is the probability of crossover, j_{rand} is a random integer in [1,n].

2.4. Selection operator

Finally, the *selection* operator is employed to maintain the most promising trial individuals in the next generation. The classical DE adopts a simple selection scheme. It compares with the objective values of the target $\vec{x}_i^g$ and trial $\vec{u}_i^g$ individuals. If the trial individual reduces the value of the objective function then it is accepted for the next generation; otherwise the target individual is retained in the population. The *selection* operator is defined as

$$\vec{x}_i^{g+1}=\begin{cases}\vec{u}_i^g, & \text{if}\quad f(\vec{u}_i^g)<f(\vec{x}_i^g)\\ \vec{x}_i^g, & \text{otherwise.}\end{cases}$$

The pseudocode of the classical DE (DE/rand/1) is illustrated in Fig. 1

```
X = initial._population (m), F, CR = initial._parameters
while ! termination_condition do
  for g = 0 to m
    v_i^g = x_{r1}^g + F(x_{r2}^g - x_{r3}^g)          // mutation
    u_i^g = binomial_crossover(x_i^g, v_i^g)          // crossover
    if f(u_i^g) ≤ f(x_i^g) then                       // selection
      x_i^g = u_i^g
    end if
    g = g + 1
  end for
end while
```

Fig. 1. Pseudocode of classical DE (DE/rand/1).