# A Clustering Based Differential Evolution

Dai-Hyun Jang, Hyung-Jin Kim, Seong-Yoon Shin*
DaiShin Information & Communications Co., Ltd, JeonBuk National University, Kunsan National University*

daijang@dsic.co.kr, kim@jbnu.ac.kr, *s3397220@kunsan.ac.kr

Abstract

We propose a novel data-driven mutation strategy for parent individuals selection, namely tensor-based DE with parapatric and cross-generation (TPCDE). Firstly, we construct a third-order population tensor to represent the relationship among generations, individuals, boundary partitions. Then the population data is classified into multiple clusters by combing tensor-based feature extraction approach and affinity propagation (AP) clustering algorithm. Finally, different parent individuals are selected from other clusters to guide the evolution.

**I. Introduction**

Big data, which are digital data with the exponential growth and wide availability, are recognized with 4Vs characteristics (volume, velocity, variety, and veracity) [1]. To retrieval potential information from big data, an effective solution for solving complex problems, has been widely used in big data. However, because of the diversity, vague and large-scale characteristics of big data, it brings unprecedented challenges for data-driven method in computational intelligence.

Ⅱ. DIFFERENTIAL EVOLUTION

This section provides some background, including DE and tensor.

## A. DE Algorithm

DE is used to solve real optimization problems. In this paper, the objective function can be expressed as (1)

$$\min f(x_1, x_2, \cdots, x_D)$$
$$s.t \quad L_i \le x_i \le U_i, \quad i=1,2,\cdots, D, \qquad (1)$$

where $f(.)$ is continuous and $xi$ is the candidate solution, D represents the dimension of the vector. $L_i$ and $U_i$ are the lower and upper bounds of $xi$.

Initially population $\{x_{i,0}=(x_{i,1,0}, x_{i,2,0}, \ldots, x_{i,D,0}),\ i=1,2,\ldots,NP\}$ is randomly generated in the search space of $x_i$, where $NP$ is the population size. Therefore, the $j$th component of $xi$ can be produced by (2).

$$x_{i,j,G}=L_i + rand_{i,j}(0,1)\cdot(U_i\text{-}L_i) \qquad (2)$$

where $G$ represents the generation number of the population evolution, $rand_{i,j}(0,1)$ is a uniformly distributed random number within the range [0, 1]. The following steps are taken next: mutation, crossover, and selection.

## B. Tensor and its Decompositions

As an extension of matrix into high-order space, tensor, has been widely used in vary domains [11], such as chemometrics, psychometrics, signal processing, data mining, and machine learning. An $m$th-order tensor may be viewed as an element of the vector space $\mathbb{R}^{n_1\times\cdots\times n_m}$. A particular component of is given by the multiply-indexed value where $i_j \in \{1,\cdots,n_j\}$ for $1 \le j \le m$.

There exist two typical tensor decomposition formats including CP decomposition [12], Tucker decomposition [13]. The tensor decomposition consists of two steps: tensor unfolding along every order and performs the SVD operation on all unfolded matrices. The n-mode matrix unfolding of tensor A are defined as: $A_n \in \mathbb{R}^{I_n\times(I_1\cdots I_{n-1}I_{n+1}\cdots I_N)}$.

## III. HE TENSOR-BASED DE ALGORITHM WITH PARAPATRIC AND CROSS-GENERATIONAL

In this section, the parapatric and cross-generational selection scheme are proposed. The first scheme is to select the parapatric individuals from the current generations; the other is to select the cross-generational elite individuals. The algorithm mainly improves the parent selection in the mutation strategy, and enhances the diversity of the population. The most important components of the algorithm will be discussed as follows as follows.

## A. Individual Distribution in DE

In DE, the population is initialized by the Eq. (2). the individual is uniformly distributed within the constrained space. During evolution, the individuals will gradually be concentrated around certain solutions. In order to observe the distribution of individuals during the evolutionary process, the distributions are depicted in Fig. 1, where the axis is the range of individual values, and the individuals connected with the same color line belong to the same group.

## B. Population Based on Tensor

In this paper, we constructed the population data into a tensor and extracted relevant information through HOSVD. Fig. 2 depicts the model of the tensor-based population. The population tensor is a fifth-order tensor namely generations, individuals, boundary partitions, strategies, and control parameters.

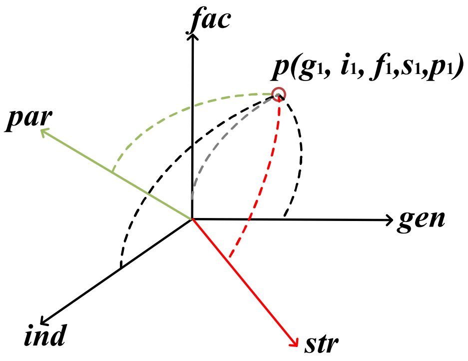


**Fig. 2.** A population tensor.

To convert the population data into high-order tensor, we need to write specific functions. Algorithm 1 presents the pseudo-code of the population tensor construction, where mainly include two parts: boundary partition and frequency calculation.

**Algorithm 1**: The population tensor construction algorithm

**Input**: the population tensor $A$, generation $g$, population $p$, partition number $k$.
**Output**: $A$ is tensor
1. Initialize the population tensor $A(g, :, :)$.
2. Calculate the *maximum* and *minimum* for each individuals in p.
3. sec ← divide the interval [*min*, *max*] into k subintervals.
4. Get the rows and columns of $p$, and assign to $r$ and $c$ respectively.
5. *len* ← the number of columns in the *sec*.
6. **for** $i \leftarrow 1$ to $r$ **do**
7. **for** $j \leftarrow 1$ to $len - 1$ **do**
8. $A(g, :, :)$ ← calculate the frequency of the $i$ individual in *sec(j)*.
9. **end for**
10. **end for**
11. Return $A$

## C. Description of Proposed Method

In this subsection, a novel tensor-based selection strategy is presented. The strategy consists of two main schemes: the parapatric selection and the cross-generation selection.

### 1) Parapatric Selection Scheme

In this paper, to increase population diversity, the individuals of mutation are selected from other groups. Firstly, population is divided into different group by AP algorithm. Then, when the $i$th individual is performing the mutation operation, groups that don't contain the $i$th individual will be composed of the set $A_i$. Finally, three different individuals indices $r_1$, $r_2$, $r_3$ are randomly selected from the set $A_i$.

According to the above analysis, the pseudo-code of the parapatric selection scheme (namely PSS) is shown in Algorithm 2. In this algorithm, $C$ represents cluster which include different individuals. Step 3 is a novel selection strategy which enhances the diversity of the population. The elements in the set $A_i$ do not contain individual $i$, and the set $A_i$ are completely different from the collection $C_i$. In step 4, parameter δ is used to control the probability of the two mechanisms being executed. The return value of the function is randomly chosen indexes $r$.

**Algorithm 2**: The Parapatric Selection Scheme

**Input**: Population size: $NP$, Groups: $C$, Threshold: δ.
**Output**: $r$ is the chosen indexes of the population.
1. **for** i ← 1 to $NP$ **do**
2. $m \leftarrow$ find($C$==$i$)
3. $A_i \leftarrow$ select individuals $C(m)$
4. **if** $rand \prec \delta$ **then**
5. Select $r_1 \neq r_2 \neq r_3 \neq i$ from the set $A_i$.
6. **else**
7. Select uniform randomly $r_1 \neq r_2 \neq r_3 \neq i$
8. **end if**
9. $r(i, :) \leftarrow r_1, r_2, r_3$.
10. **end for**
11. Return the chosen indexes $r$

### 2) Scheme Cross-Generation Selection Scheme

Adaptive Cross-generation differential evolution (ACGDE), proposed by Qiu, is one of novel mutation strategies which employs information across generations to help guide the searching directions. Unlike classic DE, ACGDE mutation strategy utilizes the information from current and previous generation to generate the donor vector. On that basis, we propose a novel cross-generational selection scheme (CGSS), which is mainly different in two aspects. First, the historical population of previous three generations has been stored into an archive. Second, the individuals of cluster centers are selected from the archive to generate the mutant vector.

To describe the CGSS selection process, an example is shown in Fig. 3, where $P$, $K$ and $N$ are three generation populations that have been stored separately after a fixed number of iterations. When the donor vectors are generated in $M$th generation, the following two steps will be executed. First, the populations of the previous three generations are clustered, and the cluster centers individuals are identified. Second, individuals are randomly selected from the cluster centers.

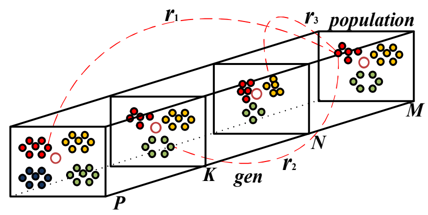


**Fig. 3.** A tensor-based cross-generation selection

# V. CONCLUSIONS

In the process of population evolution, a high order population tensor is reconstructed by the optimized population, and the potential information of population data is extracted by means of data processing. To improve the diversity of DE, the population data is divided into different groups by AP clustering algorithm, then a novel data-driven selection mechanism where the parent individuals are selected from other clusters is incorporated into the TPCDE framework. The superior performance of TPCDE is evaluated on the basis of a set of benchmark functions compared with other state-of-the-art DE variants. The experimental results show TPCDE is the effective and the efficient.

REFERENCES

[1] Yi J H, Deb S, Dong J, et al. An improved NSGA-III algorithm with adaptive mutation operator for Big Data optimization problems[J]. Future Generation Computer Systems, 2018, 88: 571-585. DOI: doi.org/10.1016/j.future.2018.06.008.