Statistic

for machine learning

Tran Trong Khiem

AI lab tranning

2024/05/29

Introduction

Differential Evolution (DE) is a novel parallel direct search:

- Population for each generation G as $\{x_{i,G}\}_0^{\mathrm{NP}-1}$
- Size of population does't change during optimization process.
- Generates new trail vector by calculate the weighted sum of three different members.
- $x_{\text{best},G}$ is evaluated for every generation G in order to keep track of the optimization progress.
- Basics scheme:
    1. Scheme DE1
    2. Scheme DE2

Scheme DE1

For each vector $x_{i,G}$, new vector v is generated according to:

$$v = x_{r_1,G} + F(x_{r_2,G} - x_{r_3,G})$$

- $r_1, r_2, r_3 \in \{0, 1, 2, \ldots, \mathrm{NP} - 1\}$, integer and mutually different.
- F is a real and constant factor.

In order to increase the diversity of the parameter vectors, the child vector $u = \left(u_1, u_2, \ldots, u_D\right)^T$.

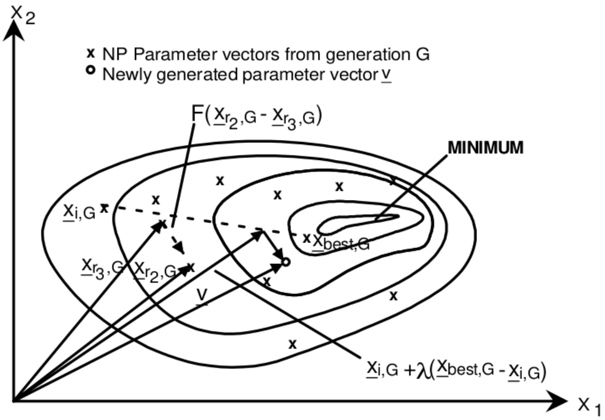
- Choose n random from $[0, D - 1]$.
- L is drawn from the interval $[0, D - 1]$ with the probability $\Pr(L = \nu) = (\mathrm{CR})^{\nu}$, where $\mathrm{CR} \in [0, 1]$.

$$u_j = \begin{cases} v_j, & \text{for } j = n \bmod D, (n+1) \bmod D, \cdots (n+L-1) \bmod D \\ (x_{i,G})_j, & \text{otherwise} \end{cases}$$

Scheme DE2

Basically, scheme DE2 works the same way as DE1 but generates the vector v according to

$$\mathbf{v} = \mathbf{x}_{i,G} + \lambda \cdot (\mathbf{x}_{\text{best},G} - \mathbf{x}_{i,G}) + F \cdot (\mathbf{x}_{r2,G} - \mathbf{x}_{r3,G})$$



JADE: Adaptive differential evolution with optional external archive

For each vector $x_{i,G}$, new vector v is generated as:

$$\mathbf{v} = \mathbf{x}_{i,G} + \lambda \cdot (\mathbf{x}_{\text{pbest},G} - \mathbf{x}_{i,G}) + F \cdot (\mathbf{x}_{r2,G} - \mathbf{x}_{r3,G})$$

x_i is associated with its own CR_i and F_i parameters $CR_i \sim \mathcal{N}(\mu_{CR}, 0.1)$ and $F_i \sim \mathcal{C}(\mu_F, 0.1)$.

- If CR_i is generated outside of the interval $[0, 1]$, it is replaced by the limit value (0 or 1) closest to the generated value.
- When $F_i > 1$, F_i is truncated to 1, and when $F_i \leq 0$, the sampling is repeatedly applied to try to generate a valid value.
- $x_{\text{pbest},G}$ is randomly selected from the top $N \times p$.

At the end of the generation,if v is better than x, then CR_i and F_i are recoded as S_{CR} and S_F, μ_{CR} and μ_F are updated as:

$$\mu_{CR} = (1 - c) \cdot \mu_{CR} + c \cdot \text{mean}_A(S_{CR})$$

$$\mu_F = (1 - c) \cdot \mu_F + c \cdot \text{mean}_L(S_F)$$

Appendix

External Archive:

- In order to to maintain diversity, if $u_{i,G}$ better than $x_{i,G}$, then $x_{i,G}$ append to archive set A.
- $x_{r2,G}$ is choosen randomly from set $P \cup A$.
- If $|A| = |P|$, randomly delete element from A.

Lehmer mean is computed as:

$$\text{mean}_L(S_F) = \frac{\sum_F F^2}{\sum_F F}$$

Weighted mean is computed as:

$$\text{mean}_W(S_{CR}) = \sum_k w_k \cdot S_{CR,k}$$

$$w_k = \frac{\Delta f_k}{\sum_k \Delta f_k}$$

SHADE:Success-History Based Parameter Adaptation for Differential Evolution

SHADE maintains a historical memory with H entries for both of the DE control parameters CR and F , M_{CR}, M_F . In each generation, for each x_i :

- Select index r_i randomly from the interval $[1, H]$.
- sample $CR_i \sim \mathcal{N}(M_{CR,r_i}, 0.1)$ and $F_i \sim \mathcal{C}(M_{C,r_i}, 0.1)$ and $p_{i,G} = \text{rand}[p_{\min}, 0.2]$
- Generate new trail vector as JADE.

If $u_{i,G}$ better than $x_{i,G}$:

- CR_i and F_i are recorded in S_{CR} and S_F.
- The contents of memory are updated as follows:

$$M_{CR,k,G+1} = \begin{cases} \text{mean}_{AW}(S_{CR}) & \text{if } S_{CR} \neq \emptyset \\ M_{CR,k,G} & \text{otherwise} \end{cases}$$

Update the same for $M_{F,k,G+1}$

1. Differential Evolution
2. JADE
3. mixSHADE

mixSHADE

Ideal: use SHADE with different with multiple mutation strategies.
Mutation strategies:

- DE/pbest/2: $v_i = x_{pbest} + F(x_{r1} - x_{r2}) + F(x_{r3} - x_{r4})$
  - Rapid convergence but reduced exploration.
- DE/current-to-pbest/1: $v_i = x_i + F(x_{r1} - x_{r2}) + F(x_{pbest} - x_i)$
  - Combines exploration and exploitation.
- DE/target-to-rand/1:
  $v_i = x_i + F(x_{r1} - x_i) + F(x_{r2} - x_i) + F(x_{r3} - x_{r2})$
  - Increases exploration by targeting random individuals.

In each generation, with each x_i, we have:

- p_1 is probability using DE/pbest/2.
- p_2 is probability using DE/current-to-pbest/1.
- p_3 is probability using DE/target-to-rand/1.

Update mutation strategies probability

Using scores to evaluate mutation strategies $\{s_1, s_2, s_3\}$.
At begin of algorithm, init $s_i = \epsilon_i$ with $\epsilon_i > 0$. Sample mutation strategies probability as:

$$p_i = \frac{|s_i|}{\sum_j |s_j|}$$

- If $u_{i,G}$ better than $x_{i,G}$ then:

$$s_i = s_i + \min(\beta_i, \frac{|f(u_i) - f(x_i)|}{|f(x_i)| + \delta})$$

  - $\beta_i > 0$ can be set different for each mutation strategies.
  - $\delta > 0$ is used to avoid division by zero.
  - Set β_i to a small value in the early generations for better exploration.

Experiment in CEC2013

Table 1: Comparison of Mean (Std Dev) across SHADE, JADE, and MixSHADE for different functions.

F	SHADE	JADE	MixSHADE
F_1	$0.00e+00$ $(0.00e+00)$	$0.00e+00$ $(0.00e+00)$	**0.00e+00 (0.00e+00)**
F_2	$9.00e+03$ $(7.47e+03)$	$7.67e+03$ $(5.66e+03)$	**7.53e+03 (7.47e+03)**
F_3	**4.02e+01 (2.13e+02)**	$4.71e+05$ $(2.35e+06)$	**4.02e+01 (2.13e+02)**
F_4	$1.92e+04$ $(3.01e+04)$	$4.10e+02$ $(4.40e+02)$	**1.90e+04 (3.01e+04)**
F_5	$0.00e+00$ $(0.00e+00)$	$0.00e+00$ $(0.00e+00)$	**0.00e+00 (0.00e+00)**
F_6	**5.96e-01 (3.73e+00)**	$2.07e+01$ $(7.17e+00)$	$6.06e-01$ $(3.73e+00)$
F_7	$4.60e+00$ $(5.39e+00)$	$2.09e+01$ $(4.93e+02)$	**4.50e+00 (5.39e+00)**
F_8	**2.07e+01 (1.76e+01)**	$2.09e+01$ $(4.93e+02)$	$2.12e+01$ $(1.76e+01)$
F_9	$2.75e+01$ $(1.77e+01)$	**2.65e+01 (1.96e+01)**	$2.71e+01$ $(1.77e+01)$
F_{10}	$7.69e-02$ $(3.58e-02)$	**4.02e-02 (2.37e-02)**	$7.53e-02$ $(3.58e-02)$
F_{11}	$0.00e+00$ $(0.00e+00)$	$0.00e+00$ $(0.00e+00)$	**0.00e+00 (0.00e+00)**
F_{12}	$2.31e+01$ $(3.73e+00)$	$2.29e+01$ $(5.45e+00)$	**2.25e+01 (3.73e+00)**
F_{13}	$5.03e+01$ $(1.34e+01)$	**4.67e+01 (1.37e+01)**	$5.01e+01$ $(1.34e+01)$
F_{14}	$3.18e-02$ $(2.33e-02)$	$2.86e-02$ $(2.53e-02)$	**2.78e-02 (2.33e-02)**

Experiment in CEC2013

Table 2: Comparison of Mean (Std Dev) across SHADE, JADE, and MixSHADE for different functions.

F	SHADE	JADE	MixSHADE
F_{15}	$3.22e+03$ $(2.64e+02)$	**1.84e+00 (6.27e-01)**	$3.16e+00(2.64e-01)$
F_{16}	**9.13e-01 (1.85e-01)**	$1.84e+00(6.27e-01)$	$9.25e-01(1.85e-01)$
F_{17}	$3.04e+01$ $(3.83e-14)$	$3.04e+01$ $(1.95e-14)$	**2.94e+01 (3.83e-14)**
F_{18}	**7.20e+01 (5.58e+00)**	$7.76e+01$ $(5.91e+00)$	$7.25e+01$ $(5.58e+00)$
F_{19}	$1.36e+00$ $(1.20e-01)$	$1.44e+00$ $(8.71e-02)$	**1.30e+00 (1.20e-01)**
F_{20}	$1.05e+01(6.04e-01)$	**1.04e+01 (5.82e-01)**	$1.07e+01(6.04e-01)$
F_{21}	$3.09e+02$ $(5.65e+01)$	$3.04e+02$ $(6.68e+01)$	**3.02e+02 (6.68e+01)**
F_{22}	**3.36e+03 (4.01e+02)**	$9.39e+01$ $(3.08e+01)$	$3.93e+03$ $(4.01e+02)$
F_{23}	$3.51e+03$ $(4.11e+02)$	**3.36e+03 (4.01e+02)**	$3.57e+03(4.12e+02)$
F_{24}	$2.05e+02$ $(5.29e+00)$	$2.17e+02$ $(1.57e+01)$	**1.85e+02 (4.29e+00)**
F_{25}	$2.59e+02$ $(1.96e+01)$	$2.74e+02$ $(1.06e+01)$	**2.53e+02 (1.97e+01)**
F_{26}	$2.02e+02$ $(1.48e+01)$	**2.15e+02 (4.11e+01)**	$2.12e+02$ $(1.4e+01)$
F_{27}	$3.88e+02$ $(1.09e+02)$	$6.70e+02$ $(2.40e+02)$	**3.78e+02 (1.e+02)**
F_{28}	$3.00e+02$ $(0.00e+00)$	$3.00e+02$ $(0.00e+00)$	$3.00e+02$ $(0.00e+00)$