

Porównanie klasycznej ewolucji różnicowej DE/rand z odmianą przesuającą punkt środkowy DE/mid

Adam Stelmaszczyk

1 listopada 2013

Streszczenie

W pracy porównano klasyczny algorytm ewolucji różnicowej DE/rand z jego odmianą DE/mid. W części teoretycznej wyprowadzono parametry skalujące dla obydwu algorytmów. W części praktycznej przeprowadzono szereg eksperymentów na 7 wybranych funkcjach z BBOB 2013 [2]. DE/mid okazał się lepszy na większości z nich.

1 Część teoretyczna

Klasyczna ewolucja różnicowa DE/rand oraz jej odmiana DE/mid różnią się jedynie operatorem mutacji. W DE/rand/ k , mutant i -tego osobnika w populacji P o n osobnikach powstaje w następujący sposób [1]:

$$u_i = P_{i_1} + \frac{F}{\sqrt{k}} \sum_{j=1}^k (P_{i_{2j}} - P_{i_{2j+1}}), i \neq i_1 \neq i_2 \neq \dots \neq i_{2k+1} \quad (1)$$

$k \in \mathbb{N}$ to liczba wektorów różnic. $F \in \mathbb{R}$ to parametr skalujący dla DE/rand/1.

W DE/mid/ k mutant powstaje w podobny sposób:

$$u'_i = m + F_m \sum_{j=1}^k (P_{i_{2j}} - P_{i_{2j+1}}), i \neq i_1 \neq i_2 \neq \dots \neq i_{2k+1}, k \in \mathbb{N} \quad (2)$$

Jedyną różnicą jest m , czyli punkt środkowy populacji:

$$m = \frac{1}{n} \sum_{j=1}^n P_j \quad (3)$$

$F_m \in \mathbb{R}$ jest parametrem skalującym dla DE/mid, analogicznym do F dla DE/rand. Żeby macierz kowariancji populacji w DE/mid/k była taka sama jak w DE/rand/k, macierz kowariancji mutantów $C[u_i]$ musi być taka sama jak macierz kowariancji mutantów $C[u'_i]$. Można to osiągnąć tak dobierając F_m , żeby było spełnione równanie:

$$C[u_i] = C[u'_i] \quad (4)$$

Osobniki są liniowo niezależne od siebie, dlatego:

$$\begin{aligned} C[u_i] &\stackrel{(1)}{=} C\left[P_{i_1} + \frac{F}{\sqrt{k}} \sum_{j=1}^k (P_{i_{2j}} - P_{i_{2j+1}})\right] = C[P_{i_1}] + C\left[\frac{F}{\sqrt{k}} \sum_{j=1}^k (P_{i_{2j}} - P_{i_{2j+1}})\right] \\ &= C[P_{i_1}] + \frac{F^2}{k} C\left[\sum_{j=1}^k (P_{i_{2j}} - P_{i_{2j+1}})\right] = C[P_{i_1}] + \frac{F^2}{k} C\left[\sum_{j=2}^{2k+1} P_{i_j}\right] \end{aligned}$$

$\forall i \ C[P_i] = C[P]$, ponieważ każdy osobnik ma taki sam rozkład prawdopodobieństwa. Zatem:

$$\begin{aligned} C[u_i] &= C[P_{i_1}] + \frac{F^2}{k} C\left[\sum_{j=2}^{2k+1} P_{i_j}\right] = C[P] + \frac{F^2}{k} C\left[\sum_{j=2}^{2k+1} P\right] \\ &= C[P] + \frac{F^2}{k} 2k C[P] = C[P] + 2F^2 C[P] = (1 + 2F^2) C[P] \end{aligned}$$

Rozwijając prawą stronę równania (4):

$$\begin{aligned} C[u'_i] &\stackrel{(2)}{=} C\left[m + F_m \sum_{j=1}^k (P_{i_{2j}} - P_{i_{2j+1}})\right] \\ &\stackrel{(3)}{=} C\left[\frac{1}{n} \sum_{j=1}^n P_j\right] + F_m^2 C\left[\sum_{j=1}^k (P_{i_{2j}} - P_{i_{2j+1}})\right] = \frac{1}{n} C[P] + F_m^2 C\left[\sum_{j=2}^{2k+1} P\right] = \left(\frac{1}{n} + 2kF_m^2\right) C[P] \end{aligned}$$

Zatem $C[u_i] = (1 + 2F^2) C[P]$ oraz $C[u'_i] = \left(\frac{1}{n} + 2kF_m^2\right) C[P]$. Podstawiając do (4):

$$(1 + 2F^2) C[P] = \left(\frac{1}{n} + 2kF_m^2\right) C[P]$$

Przy założeniu, że $C[P] \neq 0$:

$$\begin{aligned} 1 + 2F^2 &= \frac{1}{n} + 2kF_m^2 \\ \frac{1 + 2F^2 - \frac{1}{n}}{2k} &= F_m^2 \end{aligned}$$

Obie strony są nieujemne, więc:

$$F_m = \sqrt{\frac{1 + 2F^2 - \frac{1}{n}}{2k}} \quad (5)$$

Przyjmijmy $F = 0.9$. Wówczas z (5) wynika, że:

$F_m \approx 1.14$ dla $k = 1$ i $n \rightarrow \infty$.

W DE/mid/1, zamiast przesuwac losowo wybranego osobnika P_i , przesuwamy punkt srodkowy. Punkt srodkowy jest mniej zmienny, tzn. norma macierzy kowariancji punktu srodkowego jest mniejsza niz norma macierzy kowariancji dowolnego osobnika. $\lim_{n \rightarrow \infty} C[m] = 0$, natomiast $C[P_i] = C[P]$. Dlatego DE/mid/1 potrzebuje wiekszego wspolczynnika skalujacego niz DE/rand/1.

W przypadku $k \rightarrow \infty$, zgodnie z centralnym twierdzeniem granicznym, wyrazenie $\sum_{j=1}^k (P_{i_{2j}} - P_{i_{2j+1}})$ mozna traktowac jak zmienna losowa v_∞ o rozkladzie $\mathcal{N}(0, C[P])$.

Wówczas rownanie mutanta DE/rand/ ∞ mozna zapisać jako:

$$u_i = P_{i_1} + F_\infty \cdot v_\infty$$

Wyznamy F_∞ .

$$\begin{aligned} C[u_i] &= C[P_{i_1} + F_\infty \cdot v_\infty] = C[P] + 2F_\infty^2 C[P] \\ C[P] + C[F_\infty \cdot v_\infty] &= C[P] + 2F_\infty^2 C[P] \\ C[F_\infty \cdot v_\infty] &= 2F_\infty^2 C[P] \\ F_\infty^2 C[P] &= 2F_\infty^2 C[P] \\ F_\infty^2 &= 2F^2 \\ F_\infty &= \sqrt{2}F \end{aligned}$$

Rownanie mutanta DE/mid/ ∞ mozna zapisać podobnie:

$$u'_i = m + F_{\infty_m} \cdot v_\infty$$

Wyznamy F_{∞_m} .

$$\begin{aligned} C[u'_i] &= C[m + F_{\infty_m} \cdot v_\infty] = C[P] + 2F_{\infty_m}^2 C[P] \\ C[m] + C[F_{\infty_m} \cdot v_\infty] &= C[P] + 2F_{\infty_m}^2 C[P] \\ \frac{C[P]}{n} + F_{\infty_m}^2 C[P] &= C[P] + 2F_{\infty_m}^2 C[P] \\ F_{\infty_m} &= \sqrt{1 + 2F^2 - \frac{1}{n}} \end{aligned}$$

Tabela 1 podsumowuje znalezione parametry skalujące.

	parametr
DE/rand/1	F
DE/rand/k	$\sqrt{\frac{2F^2}{2k}} = \frac{F}{\sqrt{k}}$
DE/mid/k	$\sqrt{\frac{1+2F^2-\frac{1}{n}}{2k}}$
DE/rand/ ∞	$\sqrt{2F^2} = \sqrt{2}F$
DE/mid/ ∞	$\sqrt{1+2F^2-\frac{1}{n}}$

Tabela 1: Parametry skalujące

2 Część praktyczna

Zbadano 4 algorytmy:

1. DE/rand/1
2. DE/mid/1
3. DE/rand/ ∞
4. DE/mid/ ∞

Ekperymenty przeprowadzono na 7 funkcjach testowych o numerach 24, 110, 113, 116, 119, 122, 126 z BBOB 2013 [2], zaimplementowanych w języku C. Funkcje testowe są wywoływane z Javy, w której napisano algorytmy oraz procedurę testującą. Liczba wymiarów $D = 10$. Maksymalna liczba wywołań funkcji oceny $FES = 10^6$. Rozmiar populacji $n = 100$. Jeśli algorytm nie znajdował minimum, wówczas w jednym uruchomieniu, na jednej funkcji, generował 10^4 pokoleń. Na każdej funkcji algorytm był niezależnie uruchamiany 100 razy, z każdego uruchomienia zapisywany był najlepszy wynik. Parametry skalujące zgodne z tabelą 2, $F = 0.9$. Prawdopodobieństwo krzyżowania $Cr = 0.9$.

	wartość
DE/rand/1	0.9
DE/mid/1	1.14...
DE/rand/ ∞	1.27...
DE/mid/ ∞	1.62...

Tabela 2: Wartości parametrów skalujących użyte w eksperymentach

	24	110	113	116	119	122	126	suma
DE/rand/1	3	1	0	0	1	3	0	8
DE+/rand/1	3	1	1	1	1	3	3	13
DE/mid/1	3	3	3	3	3	3	1	19
DE+/mid/1	3	3	2	3	3	3	3	20

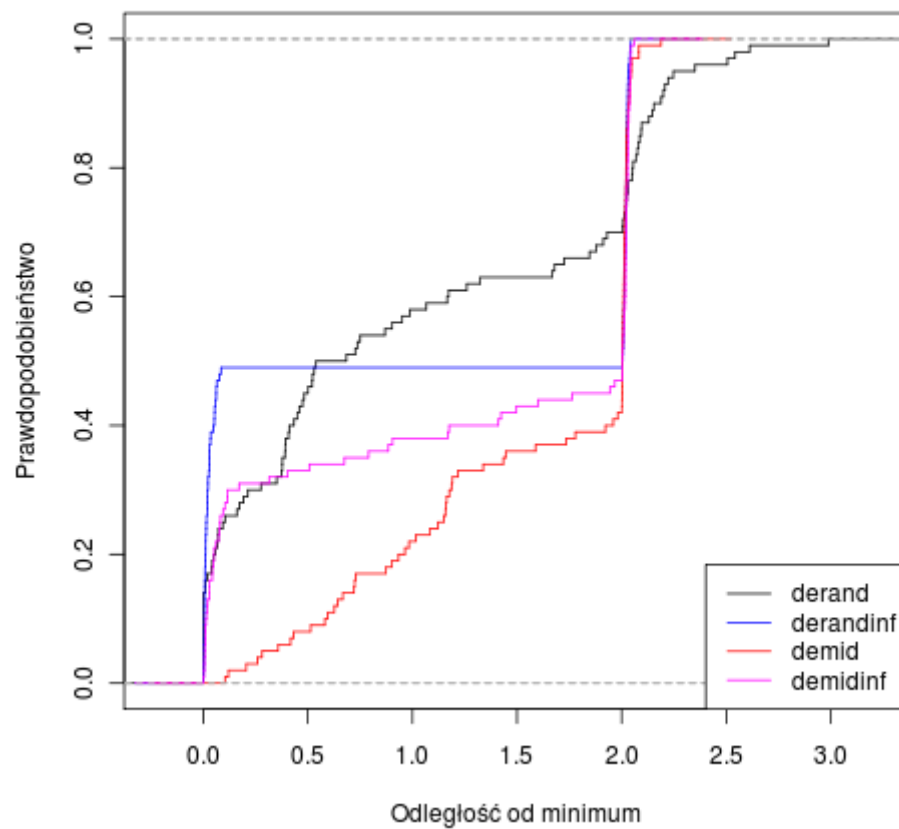
Tabela 3: Porównanie algorytmów

W celu porównania, wykreślano dystrybuanty empiryczne najlepszych wyników z każdego uruchomienia dla obu algorytmów na jednej funkcji. Wykresy przedstawiono poniżej. Najlepszym wynikiem była najmniejsza odległość funkcji oceny osobnika od minimum. Algorytm, którego dystrybuanta na wykresie przebiegała powyżej pozostałych, otrzymywał 3 punkty. Za drugie miejsce algorytm dostawał 2 punkty, za trzecie 1, za ostatnie 0. Jeśli dystrybuanty się przecinały, algorytmy zajmowały i-te miejsce ex aequo i dostawały punkty za i-te miejsce. Przykładowo, na 110 funkcji DE/mid/1 i DE+/mid/1 wygrały, a DE/rand/1 i DE+/rand/1 zajęły trzecie miejsce. Dostały następującą liczbę punktów, odpowiednio: 3, 3, 1, 1. Wyniki przedstawia tabela 3.

Para DE/mid/1 okazała się zdecydowanie lepsza niż DE/rand/1. Na 113, 116, 119, 122 i 126 funkcji DE/mid/1 znalazł minimum, podczas gdy DE/rand/1 nigdy się to nie udało. Ocenianie punktu środkowego zazwyczaj pomagało, ale nie zawsze. DE+/rand/1 nigdy nie był gorszy niż DE/rand/1. Natomiast DE+/mid/1 przegrał na 113 funkcji z DE/mid/1. Na pozostałych zazwyczaj sprawował się podobnie, raz wygrywając na funkcji 126. Funkcje 24 oraz 122 okazały się nieprzydatne, ponieważ nie wykazały jednoznacznie przewagi żadnego z algorytmów.

Referencje

- [1] Jarosław Arabas and Karol Opara. Decomposition and metaoptimization of mutation operator in differential evolution. *Swarm and Evolutionary Computation*, (LNCS 7269), 2012.
- [2] Nikolaus Hansen, Anne Auger, Steffen Finck, and Raymond Ros. Real-parameter black-box optimization benchmarking: Experimental setup. 2013. <http://coco.lri.fr/downloads/download13.09/bbobdocexperiment.pdf>.



../pngs/110.png

../pngs/113.png

../pngs/116.png

../pngs/119.png

../pngs/122.png

../pngs/126.png