УДК 519.712, 004.032.26
DOI: 10.32626/2308-5878.2023-24.81-90

Ю. А. Літвінчук
Чернівецький національний університет
імені Ю. Федьковича, м. Чернівці

САМОАДАПТИВНИЙ CMA-ES АЛГОРИТМ

У роботі буде розглянуто один із самоадаптивних алгоритмів підбору параметрів складних систем, прикладами яких являються нейронні мережі. Самоадаптивні алгоритми – це алгоритми, які змінюють свою поведінку під час виконання на основі доступної інформації та заздалегідь визначених механізмів винагороди. Ці алгоритми широко використовуються в різних галузях, включаючи машинне навчання, оптимізацію та стиснення даних. Самоадаптивність алгоритму у даному випадку буде ґрунтуватися на підборі кількості піків у суміші розподілів у розширеному CMA-ES алгоритмі за умови нормального базового розподілу.

У статті представлений покращений самоадаптивний алгоритм CMA-ES, з акцентом на параметр, що визначає кількість піків у суміші нормальних розподілів. У алгоритмі враховуються методи налаштування даного оптимального значення, що використовуються при виборі номерів кластерів в алгоритмах кластеризації CURE, BIRCH тощо. Очевидно, що наведене обґрунтування цього підходу може бути поширене на суміші з іншим базовим розподілом, кожен з яких характеризується скінченним числом піків у суміші розподілів. Це означає самоадаптивність та застосовність алгоритму до ширшого спектру сценаріїв, що включають різні характеристики розподілу. Немає сумніву у тому, що запропонований садоадаптивний алгоритм налаштування параметрів, що базується на CMA-ES алгоритмі, може бути розширений і на інші генетичні та еволюційні алгоритми які містять відбір додаткових хромосом (індивідів) при переході між ітераційними епохами алгоритму. Ще однією особливістю запропонованого алгоритму є використання теоретичних основ кластерного аналізу для оцінки кількості піків у розподілі хромосом. Даний підхід широко використовується у новітніх самоадаптивних алгоритмів для визначення початкових параметрів (гіперпараметрів) складних систем.

Ключові слова: *оптимізаційна задача, нормальний розподіл, алгоритми кластеризації, генетичний алгоритм, CMA-ES алгоритм.*

Вступ. У роботах [1-6] розглянуто різні генетичні алгоритми пошуку оптимального значення для задачі оптимізації

$$f \to MAX\ , \tag{1}$$

© Ю. А. Літвінчук, 2023

де функція $f: R^d \to R$ причому припускається що задача оптимізації є безумовною ($x \in R^d$). Важливу роль у генетичних алгоритмах відіграє метод відбору нових «особин» $x_i \in R^d$. Зрозуміло, що складність алгоритму буде залежати від розподілу «особин» на кожній із ітерацій (епох) еволюційного алгоритму. Надалі щільність розподілу нових «особин» на n -й ітерації будемо позначати через $p_{new}^{(n)}(x), x \in R^d$. Проблематику вибору початкового розподілу $p_{new}^{(0)}(x), x \in R^d$ та переобчислення $p_{new}^{(n)}(x), x \in R^d, n > 0$ детально описано в роботі [7], де увага також сконцентрована на підборі нових хромосом.

Особливу увагу в даній роботі будемо приділяти CMA-ES алгоритму (covariance matrix adaptation evolution strategy) [1,2,8-10]. Даний алгоритм ґрунтується на припущенні, що $p_{new}^{(n)}(x)$ визначається як багатовимірний нормальний розподіл, тобто

$$p_{new}^{(n)}(x; \mu_n, \Sigma_n) = (2\pi)^{-\frac{d}{2}} \left(\det(\Sigma_n)\right)^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu_n)' \Sigma_n^{-1} (x-\mu_n)},$$

де μ_n – середнє значення розподілу, Σ_n – відповідна коваріаційна матриця. Основна увага CMA-ES алгоритму приділяється саме аналізу коваріаційної матриці Σ_n , яка відображає розкид даних від свого середнього значення. Великі значення елементів коваріаційнойної матриці Σ_n вказують на великий розкид нових "особин" і в даному випадку нормальний розподіл може бути замінений рівномірним розподілом на деякій обмеженій підмножині з R^d .

Основна частина. Розглянемо один приклад використання класичного CMA-ES алгоритму для розв'язання задачі (1) та його недоліки за наявності специфічних умов функції $f \in R^d$. Розглянемо наступну оптимізаційну задачу

$$f(x) = \alpha e^{-x'x} + (1-\alpha) e^{-(x-\mu)'(x-\mu)}, x \in R^2, \qquad (2)$$

де $\alpha \in [0,1]$, $\mu = (1,1)'$. Оптимальні значення безумовної задачі оптимізації (1) для функції (2) будуть мати вигляд

$$x_{opt} = (\beta, \beta), \beta \in [0,1],$$

причому $\beta = 0$ при $\alpha = 1$ та $\beta = 1$ при $\alpha = 0$. У загальному випадку β задовольняє рівняння

$$\alpha\beta - (1-\alpha)(1-\beta) e^{4\beta-2} = 0.$$

Залежність β та $\alpha \in [0.5,1]$ бачимо на рис. 1. Використовуючи симетрію розв'язків можемо продовжити розв'язок на $\alpha \in [0,0.5]$.

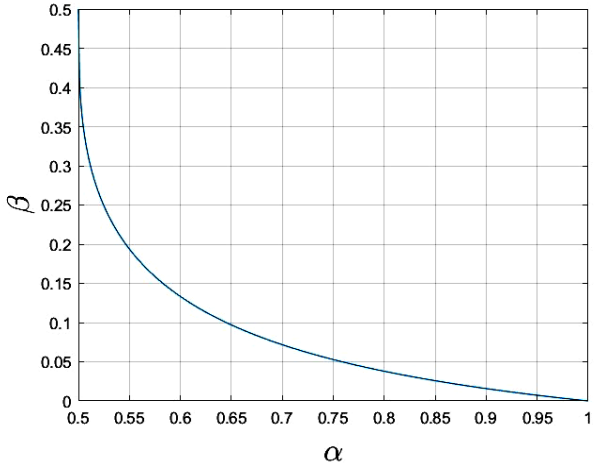

Рис. 1. *Взаємозв'язок* α *та* β *для задачі (1) та функції (2)*

Розглянемо тепер випадок використання еволюційного алгоритму CMA-ES для задачі оптимізації (1) з цільовою функцією (2). Використовуючи міркування, аналогічні до міркувань щодо оптимального розв'язку β, робимо висновок, що коваріаційна матриця відбору нових «особин» у CMA-ES алгоритмі буде мати вигляд

$$\Sigma_n = \begin{pmatrix} \sigma_1^2(n) & \rho\sigma_1(n)\sigma_2(n) \\ \rho\sigma_1(n)\sigma_2(n) & \sigma_2^2(n) \end{pmatrix},$$

де $\rho \in (0,1)$, а значення параметрів $\sigma_1(n), \sigma_2(n) > 0$ будуть залежати від параметра α. Нас буде цікавити значення $\alpha \approx \frac{1}{2}$. В цьому випадку, функція $f(x)$ (2) буде мати приблизно два піки одного рівня – рис. 2. З даного рисунка видно, що для випадку $\alpha = \frac{1}{2}$ буде мати місце рівність $\sigma_1(n) = \sigma_2(n)$ та буде справедлива нерівність

$$\sigma_1(n) = \sigma_2(n) > \sigma_c > 0, \tag{3}$$

де σ_c буде визначати нижню межу для середньоквадратичного відхилення $\sigma_1(n), \sigma_2(n)$ на кожній ітерації. Слід зауважити, що оцінка

середнього значення μ_n в CMA-ES алгоритмі буде близькою до середини між піками, тобто $\mu_n \approx \left(\frac{1}{2},\frac{1}{2}\right)$ для достатньо великих n. В загальному ж випадку μ_n буде залежати від параметра α та буде мати місце наближення $\mu_n \approx (\alpha, 1-\alpha)$. Значення граничної величини σ_c повинно бути достатньо великим для того, щоб з близькою до 1 ймовірністю на основі нормального розподілу $p_{new}^{(n)}(x;\mu_n,\Sigma_n)$ покривалися два піки функції f, тобто $(0,0)$ та $(1,1)$. Враховуючи даний факт приходимо до висновку, що $\sigma_1(n)$ та $\sigma_2(n)$ будуть лінійними функціями від відстані між піками. У випадку наявності багатьох локальних екстремумів у цільовій функції в R^2 буде мати місце наступне співвідношення

$$\sigma_1(n) = O(M), \sigma_2(n) = O(M),$$

де M визначає максимальну відстань між локальними екстремумами функції f, що є близькими до глобального екстремуму, тобто

$$M = \max_{x: f(x) \approx \max f(x)} \left| x - argmax\, f(x) \right|.$$

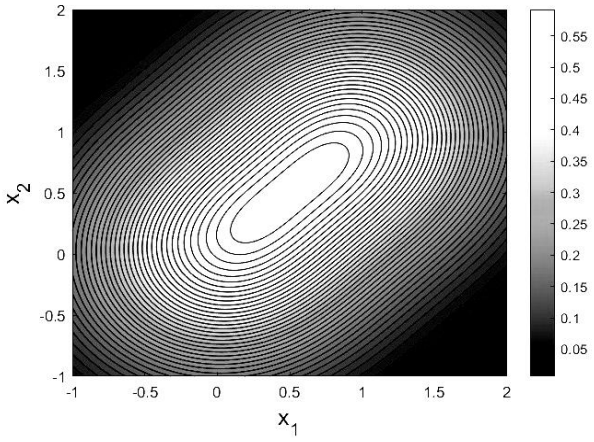

Рис. 2. *Пошуковий ландшафт для функції (2) для* $\alpha = \frac{1}{2}$

Виходячи із наведених вище міркувань приходимо до висновку, що класичний CMA-ES алгоритм із щільністю вибору на кожній ітерації $p_{new}^{(n)}(x;\mu_n,\Sigma_n)$ володіє недоліком для цільових функцій із більше ніж

одним екстремумом. У даному випадку буде суттєво збільшуватися кількість обчислень цільової функції f, що в деяких випадках має критичне значення. Для подолання цієї проблеми авторами роботи [11] було розроблено розширений CMA-ES алгоритм, який дозволяє враховувати локальні екстремуми цільової функції та локалізовувати кожен із них, враховуючи деякі особливості даного екстремуму. У роботі [11] було запропоновано заміну стандартного нормального розподілу на суміш нормальних розподілів, що визначається наступним чином

$$p_{new}^{(n)}\left(x;\mu_n,\Sigma_n,w_n,k(n)\right)= \\ =(2\pi)^{-\frac{d}{2}}\sum_{i=1}^{k(n)}w_n^i\left(\det\left(\Sigma_{n,i}\right)\right)^{-\frac{1}{2}}e^{-\frac{1}{2}\left(x-\mu_{n,i}\right)'\Sigma_{n,i}^{-1}\left(x-\mu_{n,i}\right)}, \quad (4)$$

де $\mu_{n,i}$ та $\Sigma_{n,i}$ параметри нормального розподілу в елементі суміші, $w_n=\left(w_n^1,\ldots,w_n^{k(n)}\right)$ – вагові коефіцієнти суміші, які задовольняють наступні умови

$$\begin{cases} w_n^i>0, i=1,\ldots,k(n), \\ \sum_{i=1}^{k(n)}w_n^i=1, \end{cases}$$

$k(n)$ – кількість піків у суміші (4).

Введемо до розгляду наступні позначення:

- N – загальна кількість «особин» (хромосом), що визначені в розширеному CMA-ES алгоритмі. Дане значення не змінюється при зміні ітерації *n*;
- $f_i^{(n)}=f\left(x_i\right)$ – значення цільової функції для хромосоми $x_i, i=1,\ldots,N$;
- найбільш ймовірна щільність за якою моделюється ген x_i, що визначається співвідношенням

$$a_i^{(n)}=argmax\begin{pmatrix}\left(\det\left(\Sigma_{n,1}\right)\right)^{-\frac{1}{2}}e^{-\frac{1}{2}\left(x-\mu_{n,1}\right)'\Sigma_{n,1}^{-1}\left(x-\mu_{n,1}\right)},\ldots, \\ \left(\det\left(\Sigma_{n,k(n)}\right)\right)^{-\frac{1}{2}}e^{-\frac{1}{2}\left(x-\mu_{n,k(n)}\right)'\Sigma_{n,k(n)}^{-1}\left(x-\mu_{n,k(n)}\right)}\end{pmatrix},$$

Визначення даного показника обчислюється на основі ймовірнісної кластеризації із використанням сумішей розподілів [11]: відсоток хромосом, що формуються *i*-м розподілом в суміші, тобто

$$p_i^{(n)} = \frac{\#\left\{a_j^{(n)} = i,\ j = 1,\ldots,k(n)\right\}}{k(n)}, i = 1,\ldots,k(n).$$

Алгоритм розширеного CMA-ES методу можна знайти в роботі [11]:

1. Визначення області зміни гіперпараметрів системи, розмірності суміші $k(n) = k = const$, кількості генів в генетичному алгоритмі N, точності методу ε, додаткового «параметру сталості» n_{const}.
2. $n = 1$. Задання випадковим чином початкових значень параметрів суміші (w_n, μ_n, Σ_n).
3. Вибір N генів згідно розподілу (4) з параметрами (w_n, μ_n, Σ_n) та обчислень значень цільової функції $f_i^{(n)}$.
4. Перерахунок параметрів $(w_{n+1}, \mu_{n+1}, \Sigma_{n+1})$ на основі формул ЕМ алгоритму та екстремальних значень $f_i^{(n)}$. Визначення найкращої хромосоми, що оптимізує задачу (1) та відповідного значення функції
$$F_n = \max_{i=1,\ldots,N} f_i^{(n)}.$$
5. Видалення хромосом з мінімальними значеннями $f_i^{(n)}$ та їх заміщення новими хромосомами на основі суміші (4) з параметрами $(w_{n+1}, \mu_{n+1}, \Sigma_{n+1})$.
6. Використання мутації та схрещування хромосом.
7. Якщо задовольняється умова виходу
$$\left|F_n - F_{n-n_{const}}\right| < \varepsilon$$

то оптимальний розв'язок знайдено з максимальним значенням функції F_n та хромосомою що відповідає даному оптимальному значення.

Якщо,

$$\left|F_n - F_{n-n_{const}}\right| \geq \varepsilon$$

то перейти до кроку 4.

Особливу увагу приділимо саме оцінці параметра $k(n)$, за допомогою якого вдається локалізувати оцінку локальних мінімумів функції. Основною задачею в даній статті буде визначення умов за яких буде збільшуватися чи зменшуватися кількість піків $k(n)$ на кожній ітерації.

Враховуючи дані показники, розглянемо алгоритм створення додаткового піку у суміші, тобто збільшення $k(n)$ та алгоритм вида-

лення піку, тобто зменшення $k(n)$ на одиницю. На прикладі наведеному вище можна стверджувати, що $k(n)$ повинно бути збільшити у тому випадку коли для відповідної коваріаційної матриці $\Sigma_{n,i}$ у суміші (4) відбувається затиснення до локального екстремуму, тобто

$$\lim_{n\to\infty} \det\left(\Sigma_{n,i}\right) > 0.$$

У цьому випадку пік суміші (4), що відповідає параметрам $\left(\mu_{n,i}, \Sigma_{n,i}\right)$ буде розбиватися на два піки, причому розбиття буде проводитися за допомогою класичного EM алгоритму, описаного в роботі [12]. У випадку ж видалення піку або зменшенням значення величини $k(n)$ будемо користуватися методологією алгоритмів кластеризації великих даних CURE, BIRCH [13-17]. Згідно даних алгоритмів кластерами є достатньо великі підгрупи множин, кількість елементів у яких приблизно рівна $\frac{N}{k(n)}$ або більша даного числа. В іншому випадку множина не враховується як повноцінний кластер і вважається множиною приєднання. По аналогії будемо зменшувати значення $k(n)$ у тому випадку якщо $p_i^{(n)} \ll \frac{N}{k(n)}$. Використовуючи дані алгоритми зменшення та збільшення $k(n)$, отримаємо наступний самоадаптивний розширений алгоритм CMA-ES:

1. Визначення області зміни гіперпараметрів системи, початкову розмірність суміші $k(1) = k = O\left(\ln(N)\right)$, кількості генів в генетичному алгоритмі *N*, точності методу ε, додаткового «параметру сталості» n_{const}.
2. $n = 1$. Задання випадковим чином початкових значень параметрів суміші $\left(w_n, \mu_n, \Sigma_n\right)$.
3. Вибір *N* генів згідно розподілу (4) з параметрами $\left(w_n, \mu_n, \Sigma_n\right)$ та обчислень значень цільової функції $f_i^{(n)}$.
4. Перерахунок параметрів $\left(w_{n+1}, \mu_{n+1}, \Sigma_{n+1}\right)$ на основі формул EM алгоритму та екстремальних значень $f_i^{(n)}$. Визначення найкращої хромосоми, що оптимізує задачу (1) та відповідного значення функції

$$F_n = \max_{i=1,\ldots,N} f_i^{(n)}.$$

5. Видалення хромосом з мінімальними значеннями $f_i^{(n)}$ та їх заміщення новими хромосомами на основі суміші (4) з параметрами $\left(w_{n+1}, \mu_{n+1}, \Sigma_{n+1}\right)$.
6. Використання мутації та схрещування хромосом.
7. Якщо $\det\left(\Sigma_{n,i}\right)$ не змінюються протягом n_{const} ітерацій, то збільшуємо $k(n)$ на одиницю за рахунок створення суміші з $n=2$ нормального розподілу на основі хромосом з множини $X_i = \left\{a_j^{(n)} = i,\ j = 1,\ldots,N\right\}$. Переходимо до наступного кроку.
8. Якщо $p_i^{(n)} \ll \dfrac{N}{k(n)}$ для деякого i, то видаляємо хромосоми що відповідають даному піку, тобто $X_i = \left\{a_j^{(n)} = i,\ j = 1,\ldots,N\right\}$ та зменшуємо $k(n)$ на одиницю.
9. Якщо задовольняється умова виходу

$$\left|F_n - F_{n-n_{const}}\right| < \varepsilon,$$

то оптимальний розв'язок знайдено з максимальним значенням функції F_n та хромосомою що відповідає даному оптимальному значенню.

Якщо

$$\left|F_n - F_{n-n_{const}}\right| \geq \varepsilon,$$

то переходимо до кроку 4.

Висновки. У роботі представлено розширений самоадаптивний алгоритм CMA-ES із врахуванням налаштування параметра k, що відображає кількість піків у суміші нормальних розподілів. У алгоритмі враховані методи підбору кількості кластерів, застосовані у алгоритмах кластеризації CURE та BIRCH. Немає сумніву у тому, що дана логіка може бути розвинена і на суміші із іншим базовим розподілом, які маю скінченну кількість піків.

Список використаних джерел:

1. Sakamoto N., Akimoto Y. Modified box constraint handling for the covariance matrix adaptation evolution strategy. *Proceedings of the Genetic and Evolutionary Computation Conference Companion on – GECCO '17.* 2017. URL: https://doi.org/10.1145/3067695.3075986
2. Dang V.-H., Vien N. A., Chung T. A covariance matrix adaptation evolution strategy in reproducing kernel Hilbert space. *Genetic Programming and Evolvable Machines.* 2019. URL: https://doi.org/10.1007/s10710-019-09357-1

3. Roeva O., Zoteva D., Roeva G., Lyubenova V. An Efficient Hybrid of an Ant Lion Optimizer and Genetic Algorithm for a Model Parameter Identification Problem. *Mathematics.* 2023. Vol 11. P. 1292. URL: https://doi.org/10.3390/math11061292
4. Albadr M. A., Tiun S., Ayob M., AL-Dhief F. Genetic Algorithm Based on Natural Selection Theory for Optimization Problems. *Symmetry*. 2020. Vol. 12 (11). P. 1758. URL: https://doi.org/10.3390/sym12111758
5. Xuefeng W., Chen M. Application of Mathematical Model Based on Optimization Theory and Particle Swarm Algorithm in Radar Station Layout Optimization. *Journal of Physics: Conference Series.* 2021. Vol. 1848 (1). URL: https://doi.org/10.1088/1742-6596/1848/1/012087
6. Dorsey R., Mayer W. Genetic Algorithms for Estimation Problems With Multiple Optima, Nondifferentiability, and Other Irregular Features. *Journal of Business & Economic Statistics*. 1995. Vol. 13. P. 53-66. URL: https://doi.org/10.1080/07350015.1995.10524579.
7. Alhijawi B., Awajan A. Genetic algorithms: theory, genetic operators, solutions, and applications. *Evolutionary Intelligence.* 2023. P. 1-12. URL: https://doi.org/10.1007/s12065-023-00822-6.
8. Hansen N., Müller S., Koumoutsakos P. Reducing the Time Complexity of the Derandomized Evolution Strategy with Covariance Matrix Adaptation (CMA-ES). *Evolutionary computation.* 2003. Vol. 11. P. 1-18. URL: https://doi.org/10.1162/106365603321828970.
9. Loshchilov I., Hutter F. CMA-ES for Hyperparameter Optimization of Deep Neural Networks. 2016. arXiv:1604.07269v1 [cs.NE] 25. 9 p. URL: https://doi.org/10.48550/arXiv.1604.07269
10. Hansen N., Ros R., Mauny N., Schoenauer M., Auger A. Impacts of Invariance in Search: When CMA-ES and PSO Face Ill-Conditioned and Non-Separable Problems. *Applied Soft Computing*. 2011. Vol. 11. URL: https://doi.org/10.1016/j.asoc.2011.03.001.
11. Літвінчук Ю. А., Малик І. В. Розширений алгоритм стратегії еволюції адаптації коваріаційної матриці. *Буковинський математичний журнал.* 2023. Вип. 10. № 2. С. 137-143. URL: https://doi.org/10.31861/bmj2022.02.09
12. Sundberg R. Statistical Modelling by Exponential Families. Cambridge University Press. 2019.
13. Lorbeer B., Kosareva A., Deva B., Softić D., Ruppel P., Küpper A.Variations on the Clustering Algorithm BIRCH. *Big Data Research.* 2017. Vol. 11. https://doi.org/10.1016/j.bdr.2017.09.002.
14. Lang A., Schubert E. BETULA: Numerically Stable CF-Trees for BIRCH Clustering. 2020.
15. Kogan J., Nicholas C. K., Teboulle M. Grouping multidimensional data: recent advances in clustering. Springer. 2006.
16. Guha S., Rastogi R., Shim K. CURE: An efficient clustering algorithm for large databases. *Information Systems.* 1998. Vol. 26. P. 35-58. URL: https://doi.org/10.1016/S0306-4379(01)00008-4.
17. Qian Yun-Tao, Shi Qing-Song, Wang Qi. CURE-NS: a hierarchical clustering algorithm with new shrinking scheme. 2002. Vol. 2. P. 895-899. URL: https://doi.org/10.1109/ICMLC.2002.1174512.

SELF-ADAPTIVE CMA-ES ALGORITHM

This article will consider one of the self-adaptive algorithms for selecting parameters of complex systems, examples of which are neural networks. Self-adaptive algorithms are algorithms that change their behavior at runtime based on available information and predetermined reward mechanisms. These algorithms are widely used in various fields, including machine learning, optimization, and data compression. The self-adaptiveness of the algorithm in this case will be based on the selection of the number of peaks in the mixture of distributions in the extended CMA-ES algorithm under the condition of a normal base distribution.

The work presents an improved self-adaptive CMA-ES algorithm, with an emphasis on the parameter that selects the number of pixels in a mixture of normal distributions. The algorithm takes into account the methods of setting this optimal value, which is used when choosing cluster numbers in the CURE, BIRCH, etc. clustering algorithms. It is obvious that the given justification of this approach can be extended to mixtures with a different base distribution, each of which is characterized by a skin number of peaks in the mixture distribution. This implies self-adaptability and applicability of the algorithm to a wider range of scenarios involving different distribution characteristics.

There is no doubt that the proposed sado-adaptive parameter setting algorithm, based on the CMA-ES algorithm, can be extended to other genetic and evolutionary algorithms that include the selection of additional chromosomes (individuals) during the transition between iteration epochs of the algorithm. Another feature of the proposed algorithm is the use of theoretical foundations of cluster analysis to estimate the number of peaks in the distribution of chromosomes. This approach is widely used in the latest self-adaptive algorithms for determining the initial parameters (hyperparameters) of complex systems.

Keywords: *optimization problem, normal distribution, clustering algorithms, genetic algorithm, CMA-ES algorithm.*

Отримано: 22.11.2023