

Малик І.В. Літвінчук Ю.А.

ПРО ОДИН ПІДХІД ПОБУДОВИ САМОАДАПТИВНИХ АЛГОРИТМІВ НА ОСНОВІ СУМІШЕЙ РОЗПОДІЛУ

Самоадаптивні алгоритми на даний момент є найбільш досліджуваними алгоритмами, як в теорії алгоритмів в цілому, так і в машинному навчанні, глибинному навчанні та штучному інтелекті. Активність щодо вивчення даних алгоритмів поряд із вивченням самоадаптивних нейронних мереж диктується широким спектром використання автоматичних або автоматизованих систем, які розв'язують різні задачі машинного навчання, зокрема класифікації та відновлення регресії. З точки зору класичної теорії вибору топології нейронної мережі, задача зводиться до знаходження алгоритму із автоматичного підбору гіперпараметрів нейронної мережі, а саме, кількості прихованих шарів, розмірності шарів, функцій активації, тощо.

У даній роботі вказано один із підходів, що дозволить будувати самоадаптивні алгоритми оцінки параметрів (гіперпараметрів) складних систем та узагальнювати класичні генетичні та еволюційні алгоритми. Основна ідея, яку буде висвітлено в даній роботі, базується на припущенні про мультимодальність цільової функції та ефективність використання сумішей розподілів замість одномодальних розподілів у класичному випадку. Однією із головних проблем у даній роботі є оцінка розмірності суміші та алгоритми збільшення чи зменшення суміші. Методи збільшення чи зменшення розмірності суміші базуються на методах кластерного аналізу, які використовуються в алгоритмах кластеризації великих даних CURE та BIRCH. Аналіз самоадаптивного алгоритму на основі суміші розподілів розглянемо на прикладі CMA-ES алгоритму, який є одним із основних представників алгоритмів з одномодальними розподілами нових хромосом у генетичному алгоритмі. Вказаний у роботі підхід може бути використаний і для інших оптимізаційних алгоритмів класифікації чи відновлення регресії.

Ключові слова і фрази: суміш розподілів, нормальний розподіл, алгоритми кластеризації, оптимізаційна задача, генетичний алгоритм, CMA-ES алгоритм.

Yuriy Fedkovych Chernivtsi National University, Chernivtsi, Ukraine
e-mail: j.litvinchuk@chnu.edu.ua, i.malyk@chnu.edu.ua

ВСТУП

Роботи [1, 2, 3, 4, 5, 6, 7, 9] присвячені розгляду генетичних та еволюційних алгоритмів для розв'язання різного роду оптимізаційних задач. Слід зауважити, що використання евристичних (генетичних чи еволюційних алгоритмів) зумовлено неможливістю

УДК 004.021, 519.712, 004.032.26

2010 *Mathematics Subject Classification:* 30E10, 62F10, 65E05, 68T05.

використання класичних градієнтних алгоритмів, таких, як метод Ньютона чи його модифікацій. Отже, розглянемо оптимізаційну задачу

$$f(x) \rightarrow \text{extr}, \quad (1)$$

де функція $f : R^d \rightarrow R$, причому припускається що задача оптимізації (1) є безумовною, тобто ($x \in R^d$). В сонові генетичних та еволюційних алгоритмів основну роль відграють функції схрещування, мутації та відбору нових хромосом. Зосередимо свою увагу саме на розподілі нових хромосом, причому будемо вважати, що розподіл вибору нових хромосом буде залежати від ітерації (епохи) n , на якій буде використувуватися даний розподіл. Позначимо через $p_{new}^{(n)}(x)$, $x \in R^d$, щільність розподілу хромосом на n -ій ітерації. В СМА-ES (Covariance matrix adaptation evolution strategy – Коваріаційна матриця стратегії еволюції адаптації) алгоритмі [1, 2, 8, 10] припускається що $p_{new}^{(n)}$ відповідає нормальному розподілу, тобто

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

де μ_n – середнє значення розподілу, Σ_n – відповідна коваріаційна матриця. Основна увага СМА-ES алгоритму приділяється саме аналізу коваріаційної матриці Σ_n , яка відображає розкид даних від свого середнього значення.

Слід зауважити, що у випадку мультимодальної функції класичний СМА-ES алгоритм стикається із великими значеннями елементів коваріаційної матриці Σ_n , тобто відбір нових хромосом буде здійснено не найоптимальнішим способом. Крім того, при наявності екстремумів приблизно одного рівня, з високою ймовірністю (що відповідає щільності $p_{new}^{(n)}(x; \mu_n, \Sigma_n)$) всі екстремуми повинні бути близько середнього значення.

ПРО ОДИН ПІДХІД ПОВБУДОВИ САМОАДАПТИВНИХ АЛГОРИТМІВ НА ОСНОВІ СУМІШЕЙ РОЗПОДІЛУ

Виходячи із наведених вище міркувань приходимо до висновку, що класичний СМА-ES алгоритм із щільністю вибору на кожній ітерації $p_{new}^{(n)}(x; \mu_n, \Sigma_n)$ – володіє недоліком збільшенням кількості обчислень 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 СМА-ES алгоритм, який дозволяє враховувати локальні екстремуми цільової функції та локалізувати кожен із них, враховуючи деякі особливості заданого екстремуму. У роботі [11] було запропоновано заміну стандартного нормального розподілу на суміш нормальних розподілів, що визначається наступним чином

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

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

$$\begin{cases} w_n^i > 0, \quad i = 1, \dots, k(n), \\ \sum_{i=1}^{k(n)} w_n^i = 1, \end{cases} \quad (3)$$

де $k(n)$ - кількість піків у суміші (3). Введемо до розгляду наступні позначення:

1. N – загальна кількість “особин” (хромосом), що визначені в розширеному СМА-ES алгоритмі. Дане значення не змінюється при зміні ітерації n .
2. $f_i^{(n)} = f(x_i)$ – значення цільової функції для хромосоми x_i , $i = 1, \dots, N$.
3. Найбільш ймовірна щільність за якою моделюється хромосома x_i , що визначається співвідношенням

$$a_i^{(n)} = \operatorname{argmax} \left(\begin{array}{c} (\det(\Sigma_{n,1}))^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu_{n,1})' \Sigma_{n,1}^{-1} (x-\mu_{n,1})}, \dots, \\ (\det(\Sigma_{n,k(n)}))^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu_{n,k(n)})' \Sigma_{n,k(n)}^{-1} (x-\mu_{n,k(n)})} \end{array} \right).$$

Визначення даного показника обчислюється на основі ймовірнісної кластеризації із використанням сумішей розподілів [11].

4. Відсоток хромосом, що формуються i -м розподілом в суміші, тобто

$$p_i^{(n)} = \frac{\#\{a_j^{(n)} = i, j = 1, \dots, k(n)\}}{k(n)}, \quad i = 1, \dots, k(n).$$

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

1. Визначення області зміни гіперпараметрів системи, розмірності суміші $k(n) = k = \operatorname{const}$, 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 з розподілом (3) 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 (Expectation-maximization) алгоритму та екстремальних значень $f_i^{(n)}$. Визначення найкращої хромосоми, що оптимізує задачу (1) та відповідного значення функції

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

5. Видалення хромосом з мінімальними значеннями $f_i^{(n)}$ та їх заміщення новими хромосомами на основі суміші (3) з параметрами $(w_{n+1}, \mu_{n+1}, \Sigma_{n+1})$.

6. Використання мутації та схрещування хромосом.
7. Якщо задовольняється умова виходу

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

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

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

то перейти до кроку 4.

Зосередимо увагу на оцінці параметру $k(n)$, який імітує кількість локальних екстремумів цільової функції f .

Враховуючи дані показники, розглянемо алгоритм створення додаткового піку у суміші, тобто збільшення $k(n)$ та алгоритм видалення піку, тобто зменшення $k(n)$ на одиницю. Зрозуміло, що $k(n)$ повинно бути збільшене у тому випадку коли для відповідної коваріаційної матриці $\Sigma_{n,i}$ у суміші (3) відбувається стиснення до локального екстремуму, тобто

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

В цьому випадку пік суміші (3), що відповідає параметрам

$$(\mu_{n,i}, \Sigma_{n,i})$$

буде розбиватися на два піки, причому розбиття буде проводитися за допомогою класичного EM алгоритму, описаного в роботі [12], причому вибір кількості піків буде аналогічним вибору кількості кластерів в CURE (Clustering using representatives), BIRCH (Balanced iterative reducing and clustering using hierarchies) [13, 14, 15, 16, 17]. На основі даних алгоритмів великі кластери з прогалинами повинні бути розбиті на менші підкластери, і реалізація цього розбиття відбувається на основі використання підвибірок. За аналогією будемо зменшувати значення $k(n)$ у тому випадку, якщо $p_i^{(n)} \ll \frac{N}{k(n)}$. Використовуючи дані алгоритми зменшення та збільшення $k(n)$, отримуємо наступний самоадаптивний розширений алгоритм СМА-ES:

1. Визначення області зміни гіперпараметрів системи, початкову розмірність суміші $k(1) = k = O(\ln(N))$, 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 з розподілом (3) 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

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

5. Видалення хромосом з мінімальними значеннями $f_i^{(n)}$ та їх заміщення новими хромосомами на основі суміші (3) з параметрами $(w_{n+1}, \mu_{n+1}, \Sigma_{n+1})$.
6. Використання мутації та схрещування хромосом.
7. Якщо

$$\det(\Sigma_{n,i})$$

не змінюються протягом n_{const} ітерацій, то збільшуємо $k(n)$ на одиницю за рахунок створення суміші з $n = 2$ з нормального розподілу на основі хромосом з множини

$$X_i = \left\{ a_j^{(n)} = i, j = 1, \dots, N \right\} .$$

8. Якщо $p_i^{(n)} \ll \frac{N}{k(n)}$ для деякого i , то видаляємо хромосоми, що відповідають даному піку, тобто $X_i = \left\{ a_j^{(n)} = i, j = 1, \dots, N \right\}$ та зменшуємо $k(n)$ на одиницю.
9. Якщо задовольняється умова виходу

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

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

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

то переходимо до кроку 4.

ВИСНОВКИ

Робота присвячена висвітленню нової методології побудови самоадаптивних алгоритмів на прикладі використання узагальненого СМА-ES алгоритму. У роботі розглянуто умови збільшення та зменшення розмірності суміші, що є визначальною для розширеного СМА-ES алгоритму [11]. За основу вибору оптимальної (субоптимальної) розмірності суміші (2) розглянуто методологію визначення оптимальної кількості кластерів у алгоритмах кластеризації великих даних, а саме CURE та BIRCH. Розроблена методологія може бути розширена на інші генетичні та еволюційні алгоритми, особливістю яких є вибір нових хромосом на кожній із ітерацій. Особливу увагу в роботі приділено вибору розмірності суміші на основі аналогічних задач кластерного аналізу, що дозволяє використати методи кластерного аналізу в задачах оптимізації для мультимодальних цільових функцій.

СПИСОК ЛІТЕРАТУРИ

- [1] Sakamoto, N., & Akimoto, Y. (2017). Modified box constraint handling for the covariance matrix adaptation evolution strategy. Proceedings of the Genetic and Evolutionary Computation Conference Companion on - GECCO '17. <https://doi.org/10.1145/3067695.3075986>
- [2] Dang, V.-H., Vien, N. A., & Chung, T. (2019). A covariance matrix adaptation evolution strategy in reproducing kernel Hilbert space. Genetic Programming and Evolvable Machines. <https://doi.org/10.1007/s10710-019-09357-1>
- [3] Roeva, O.; Zoteva, D.; Roeva, G.; Lyubanova, V. (2023). An Efficient Hybrid of an Ant Lion Optimizer and a Genetic Algorithm for a Model Parameter Identification Problem. *Mathematics*, 11, 1292. <https://doi.org/10.3390/math11061292>
- [4] Albadr, M. A., Tiun, S., Ayob, M., & AL-Dhief, F. (2020). Genetic Algorithm Based on Natural Selection Theory for Optimization Problems. *Symmetry*, 12(11), 1758. <https://doi.org/10.3390/sym12111758>
- [5] Xuefeng, W., & Chen, M. (2021). Application of Mathematical Model Based on Optimization Theory and Particle Swarm Algorithm in Radar Station Layout Optimization. *Journal of Physics: Conference Series*, 1848(1), 012087. <https://doi.org/10.1088/1742-6596/1848/1/012087>
- [6] Dorsey, Robert & Mayer, Walter. (1995). Genetic Algorithms for Estimation Problems With Multiple Optima, Nondifferentiability, and Other Irregular Features. *Journal of Business & Economic Statistics*. 13. 53-66. <https://doi.org/10.1080/07350015.1995.10524579>.
- [7] Alhijawi, Bushra & Awajan, Arafat. (2023). Genetic algorithms: theory, genetic operators, solutions, and applications. *Evolutionary Intelligence*. 1-12. [10.1007/s12065-023-00822-6](https://doi.org/10.1007/s12065-023-00822-6).
- [8] Hansen, Nikolaus & Müller, Sibylle & Koumoutsakos, Petros. (2003). Reducing the Time Complexity of the Derandomized Evolution Strategy with Covariance Matrix Adaptation (CMA-ES). *Evolutionary computation*. 11. 1-18. <https://doi.org/10.1162/106365603321828970>.
- [9] Ilya Loshchilov, Frank Hutter. CMA-ES for Hyperparameter Optimization of Deep Neural Networks. (2016). arXiv:1604.07269v1 [cs.NE] 25. – 9p. <https://doi.org/10.48550/arXiv.1604.07269>
- [10] Hansen, Nikolaus & Ros, Raymond & Mauny, Nicolas & Schoenauer, Marc & Auger, Anne. (2011). Impacts of Invariance in Search: When CMA-ES and PSO Face Ill-Conditioned and Non-Separable Problems. *Applied Soft Computing*. 11. <https://doi.org/10.1016/j.asoc.2011.03.001>.
- [11] Malyk I.V., Litvinchuk Yu.A. (2023). The extended CMA-ES algorithm. *Bukovinian Math Journal*. 10, 2, 137-143. <https://doi.org/10.31861/bmj2022.02.09>.
- [12] Sundberg, Rolf (2019). *Statistical Modelling by Exponential Families*. Cambridge University Press.
- [13] Lorbeer, Boris & Kosareva, Ana & Deva, Bersant & Softić, Dženan & Ruppel, Peter & Küpper, Axel. (2017). Variations on the Clustering Algorithm BIRCH. *Big Data Research*. 11. [10.1016/j.bdr.2017.09.002](https://doi.org/10.1016/j.bdr.2017.09.002).
- [14] Lang, Andreas & Schubert, Erich. (2020). BETULA: Numerically Stable CF-Trees for BIRCH Clustering.
- [15] Kogan, Jacob; Nicholas, Charles K.; Teboulle, M. (2006). *Grouping multidimensional data: recent advances in clustering*. Springer.
- [16] Guha, Sudipto & Rastogi, Rajeev & Shim, Kyuseok. (1998). CURE: An efficient clustering algorithm for large databases. *Information Systems*. 26. 35-58. [10.1016/S0306-4379\(1\)00008-4](https://doi.org/10.1016/S0306-4379(1)00008-4).
- [17] Qian, Yun-Tao & Shi, Qing-Song & Wang, Qi. (2002). CURE-NS: a hierarchical clustering algorithm with new shrinking scheme. 2. 895 - 899 vol.2. [10.1109/ICMLC.2002.1174512](https://doi.org/10.1109/ICMLC.2002.1174512).

Надійшло 08.12.2023

Malyk I.V., Litvinchuk Y.A. *About one approach to the construction of self-adaptive algorithms based on distribution mixtures*, Bukovinian Math. Journal. **11**, 2 (2023), 183–189.

This article presents a novel approach for constructing self-optimizing algorithms designed to estimate parameters (hyperparameters) in complex systems, with a broader application to classical genetic and evolutionary algorithms. The central theme of this paper revolves around the exploration of multimodality in the objective function and advocates the effectiveness of employing distribution mixtures as opposed to single-peaked distributions in traditional scenarios. A significant focus of this research involves addressing the challenge of determining the dimensionality of the mixture and developing algorithms for both augmenting and reducing it. The methods employed for manipulating the mixture's dimensionality are inspired by cluster analysis techniques, specifically those utilized in the CURE and BIRCH big data clustering algorithms. Furthermore, this work delves into a detailed examination of a self-adaptive algorithm grounded in a mixture of distributions, illustrated by the CMA-ES algorithm. It is evident that the proposed approach outlined in this paper exhibits versatility, making it applicable not only to the CMA-ES algorithm but also to various optimization algorithms involved in tasks such as classification or regression recovery.