EVOLUTIONARY ALGORITHMS FOR SOLVING COMPLEX OPTIMIZATION PROBLEMS IN DATA SCIENCE

Galhenage Gayan Sudesh Suranga Perera
Enroll no: KUSLS20220143119

Abstract: *This study presents a hybrid multi-evolutionary algorithm that integrates the strengths of Genetic Algorithms (GAs) and Evolutionary Strategies (ESs) to effectively tackle complex optimization problems in data science. The proposed GA-ES approach leverages the exploratory capabilities of GAs alongside the rapid convergence attributes of ESs, facilitating a balanced navigation of diverse solution landscapes. Experimental results demonstrate that this hybrid methodology outperforms traditional optimization techniques, particularly in scenarios characterized by multiple local optima. Additionally, the interchange of top-performing individuals between the two algorithms enhances optimization efficiency and leads to superior solutions within a shorter computational timeframe. The findings highlight the potential of evolutionary algorithms as robust tools for addressing intricate optimization challenges. Future research is encouraged to refine the hybrid framework and investigate its application across a wider array of real-world data science problems, particularly those necessitating adaptive optimization strategies, thereby contributing to advancements in data-driven decision-making processes.*

Keywords: *Hybrid Multi-Evolutionary Algorithm, Genetic Algorithms, Evolutionary Strategies, Functions Optimization Problem, Objective Function Optimization.*

1. Introduction

In the era of data-driven decision-making, the ability to solve complex optimization problems has become increasingly critical across various fields, including finance, engineering, and artificial intelligence. Traditional optimization methods often struggle with the intricacies and high-dimensional nature of these problems, particularly when faced with multiple local optima and diverse solution landscapes. Evolutionary algorithms, such as Genetic Algorithms (GAs) and Evolutionary Strategies (ESs), have emerged as powerful alternatives due to their adaptive nature and capability to explore vast search spaces. This study introduces a hybrid multi-evolutionary algorithm that integrates the strengths of GAs and ESs to address the challenges posed by complex optimization tasks in data science. By combining the exploratory capabilities of GAs, which excel in diverse solution searches, with the rapid convergence properties of ESs, this novel GA-ES approach enhances optimization efficiency and effectiveness. Preliminary experimental results indicate that the hybrid methodology significantly outperforms traditional techniques, particularly in scenarios characterized by numerous local optima. The ability to interchange top-performing individuals between GAs and ESs not only boosts optimization efficiency but also fosters the discovery of superior solutions in reduced computational timeframes. This innovative framework positions evolutionary algorithms as robust tools for tackling intricate optimization challenges. The implications of this work extend to a wide array of real-world applications, encouraging further research to refine the hybrid model and explore its potential in adaptive optimization strategies across diverse data science problems.

2. Problem Statement

Traditional optimization methods struggle with complex, high-dimensional problems in data-driven decision-making. This study introduces a hybrid multi-evolutionary algorithm combining Genetic Algorithms and Evolutionary Strategies, enhancing optimization efficiency and effectiveness to address these challenges in diverse real-world applications.

3. Literature Review

Evolutionary algorithms (EAs) have emerged as powerful tools for addressing complex optimization problems in data science. Inspired by natural selection, these algorithms effectively explore vast solution spaces, adapting to intricate problem landscapes. Their versatility allows application across various domains, including machine learning, feature selection, and resource allocation. As data complexity increases, the need for efficient optimization methods becomes paramount, prompting researchers to investigate and refine EAs. This literature review examines the advancements, challenges, and emerging trends in EAs for data science applications.

Summary of Literature Survey

Authors	Work Done	Findings
Yu et al., 2021	Modeled COVID-19 molecular mechanisms induced by cytokine storms during SARS-CoV-2 infection	Found significant insights into COVID-19's molecular mechanism, helping understand cytokine storm effects
Yu et al., 2021	Used self-organizing maps to analyze COVID-19 SEIRS delayed model	Provided a parametric analysis, contributing to improved epidemic modeling
AbuJarad et al., 2020	Conducted Bayesian reliability analysis of Marshall and Olkin model	Enhanced the reliability analysis using Bayesian techniques, improving model precision
Sohail et al., 2020	Developed computational framework to explore environmental stress impact on epidemics	Demonstrated the effects of environmental factors on epidemic dynamics
Yu et al., 2020	Created a delayed model to forecast the periodic behavior of SARS-CoV-2	Provided better prediction methods for understanding SARS-CoV-2 behavior
Sun et al., 2020	Designed CNN architectures for image classification using a genetic algorithm	Improved image classification accuracy through the genetic algorithm-generated CNN architectures
Lin et al., 2020	Proposed a new optimization model of CCHP system based on genetic algorithm	Enhanced energy system efficiency with a genetic algorithm-based optimization model
Drezner & Drezner,	Developed biologically inspired parent selection methods in genetic algorithms	Improved the performance of genetic algorithms with biologically inspired selection techniques

2020		
Gramaje et al., 2019	Applied machine learning techniques for patient discharge classification	Improved hospital discharge decision-making processes using machine learning methods
Sohail, 2019	Applied Bayesian machine learning for biomedical data set inference	Improved data set inference accuracy for biomedical applications using Bayesian methods
Iftikhar et al., 2019	Performed deterministic and stochastic analysis of a dengue spread model	Provided enhanced epidemic modeling techniques, improving predictions for dengue spread
Chen et al., 2019	Developed an improved genetic algorithm with back-propagation neural network for water-level predictions	Enhanced accuracy of water-level predictions by combining genetic algorithm and neural network techniques

4. Research Gap

Despite significant advancements in evolutionary algorithms (EAs) for solving complex optimization problems in data science, several research gaps remain. Many existing EAs struggle with balancing exploration and exploitation, leading to premature convergence or suboptimal solutions. Additionally, scalability issues arise when applied to high-dimensional data, which limits their effectiveness in real-world applications. Furthermore, hybrid models combining EAs with other optimization techniques are still underexplored, and more research is needed on adaptive and automated algorithm selection tailored to specific data science problems. Addressing these gaps could enhance EA performance in diverse applications.

5. Methodology

The methodology for addressing the Functions Optimization Problem (FOP) employs a hybrid multi-evolutionary algorithm that combines the strengths of Genetic Algorithms (GAs) and Evolutionary Strategies (ESs). This approach is designed to optimize decision variables within specified constraints, aiming for either maximum or minimum values of an objective function. The algorithm begins by exploring the solution space through GA's crossover and mutation operations, complemented by ES's rapid convergence capabilities. Various evolutionary strategies, including 1+1 ES and $\mu+\lambda$ ES, were evaluated in two computational stages using diverse functions, ranging from simple to highly complex, to identify optimal strategies. Performance metrics, such as average running time and fitness function calls, were analyzed to assess efficiency. Results indicated that the hybrid GA-ES algorithm outperformed traditional methods, showcasing its effectiveness in solving both continuous and discrete optimization challenges in data science. This integrated approach holds promise for enhancing the accuracy and speed of optimization tasks.

6. Result Discussion

Problem Formulation: Optimization involves selecting the optimal element from a set based on specific criteria while adhering to certain constraints. The Functions Optimization Problem (FOP) requires adjusting the parameters (variables) of a function to attain the best possible value—either maximal or minimal—of an objective function, subject to defined constraints. For example, a function maximization problem can be articulated as follows:

$$\begin{cases} \max f(x) \\ subject\ to: c_i \leq 0\ for\ i = 1, 2, \ldots, k \\ x \in S \end{cases} \tag{1}$$

where $x = [x_1, x_2, x_3, \ldots, x_n] \in R^n$, $n \in N$ represents an n-dimensional vector of decision variables, f(x) denotes the objective function of the variables x, $c_i(x)$ are the constraints, and S is the search space. The Functions Optimization Problem (FOP) can serve as a benchmark for various optimization techniques, including Genetic Algorithms and Evolutionary Strategies. Several methods for solving the FOP have been discussed, particularly

The Proposed Hybrid Multi-Evolutionary Algorithm: Genetic Algorithms (GAs) are capable of exploring a solution space through crossover and mutation operators to address complex optimization problems. While GAs typically use binary representations for individuals, they can also work with real numbers or more intricate data structures. The effectiveness of GAs is significantly influenced by the initial population, making them particularly well-suited for discrete problems. However, one limitation is their reduced efficiency during the final stages of the search process. In contrast, Evolutionary Strategies (ESs) primarily utilize real numbers to represent individuals and focus on mutation and selection as their main evolutionary operators. The competitive selection between parents and offspring creates evolutionary pressure, but ESs can sometimes become trapped in sub-optimal solutions. Their strength lies in their ability to quickly converge to optimal values, making them ideal for continuous optimization problems.

Combining the strengths of both algorithms through hybridization presents a promising approach for solving a variety of continuous and discrete optimization challenges. The proposed algorithm, GA-ES, integrates the GA's exploration capabilities to uncover potential optimal areas with the rapid convergence characteristics of ESs. Both GAs and ESs can utilize the same selection, mutation, and individual representation operators, starting from a shared initial population. After a set number of generations, the best individuals from both algorithms are compared. Based on the results of this comparison, individuals may be exchanged between the algorithms:

- If the best individual from the GA outperforms the best individual from the ES, it indicates a discovery of a new optimal area. The best individual from the GA will then replace a parent in the ES, or the least fit parent if there are multiple parents.
- Conversely, if the best individual from the ES surpasses the GA's best individual, it signifies that the ES has found a new optimal solution. In this case, the best individual from the ES will replace the least fit individual in the GA population.

Computational Experiment: The proposed algorithm can be applied to various optimization problems, including the Function Optimization Problem. The experiments were conducted in two stages. In the first stage, different types of evolutionary strategies were compared, such as *1+ 1 ES, μ + 1 ES, μ, λ ES*, and *μ + λ ES.* The objective of this stage was to identify the optimal type of Evolutionary Strategy that is most effective for solving the test problems.

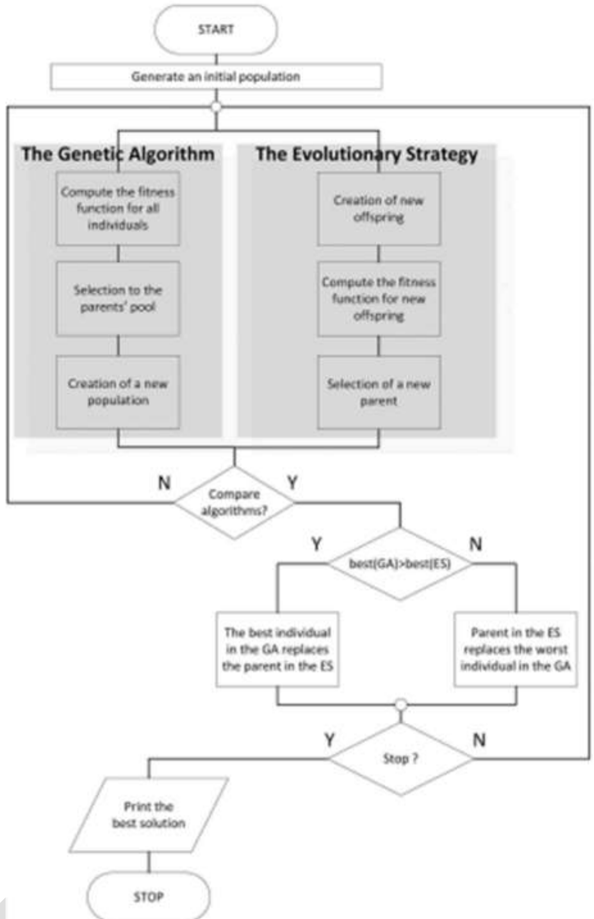

Figure 1 The hybrid algorithm block diagram.

A diverse set of functions with varying complexities was utilized in the research, including: $f_1(x_1, x_2)$: This is a simple function of two variables characterized by multiple local optima. It was employed to assess the algorithm's capability to identify the global optimum. The function is defined by the following formula:

$$f_1(x_1, x_2) = (sin(x_1) + 0.6 * sin(20 * x_1)) * sin(x_2) \quad (2)$$

where: $x_1; x_2 \in (0, \pi)$

The value of maximum 1.6 at point ($\pi/2, \pi/2$).

$f_2(x_1, x_2, \ldots, x_{10})$: This function involves multiple variables and features a low inclination angle. It was utilized to evaluate the algorithm's capacity to pinpoint the exact solution. The function is defined by the following formula:

$$f_2(x_1, x_2, \ldots, x_{10}) = \prod_{i=1}^{10} \sin(x_i) \quad (3)$$

where: $x_1; x_2 \in (0, \pi)$

The value of maximum 1 at point ($\frac{\pi}{2}, \frac{\pi}{2}, \frac{\pi}{2}, \frac{\pi}{2}, \frac{\pi}{2}, \frac{\pi}{2}, \frac{\pi}{2}, \frac{\pi}{2}, \frac{\pi}{2}, \frac{\pi}{2}$).

f_3: This function is highly complex and features multiple local optima with varying values. It is designed to evaluate the algorithm's effectiveness in tackling challenging optimization problems. In this case, the initial generation of individuals is positioned at a local optimum located at point [5,5] with a value of 1.5. The algorithm is tasked with identifying the global optimum at point [50,50] with a value of 2.5, while navigating around local optima with values of 1.0 and 2.0. The function is defined by the following formula:

$$f_3(x_1, x_2) = \sum_{1}^{7} h_i * e^{-\mu_i*(x_1-x_{1_i})^2+(x_2-x_{2_i})^2}$$

$$\text{where: } h_1 = 1.5, h_2 = 1, h_3 = 1, h_4 = 1, h_5 = 2, h_6 = 2, h_7 = 2.5$$

$$\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7 = 0.01$$

$$(x_{11}, x_{12}) = (5,5), (x_{21}, x_{22}) = (5,30), (x_{31}, x_{32}) = (25,25),$$

$$(x_{41}, x_{42}) = (30,5), (x_{51}, x_{52}) = (50,20), (x_{61}, x_{62}) = (20,50),$$

$$(x_{71}, x_{72}) = (50,50) \quad (4)$$

Rastrigin Function: This is a non-convex, non-linear multimodal function that was initially introduced as a two-dimensional function and has since been generalized to an n-dimensional domain, with experiments conducted in 2, 5, and 10 dimensions. Locating the minimum of this function is particularly challenging due to the vast search space and the numerous local minima present. The function is typically evaluated over the hypercube defined by $x\ i \in [-5.12,5.12]$ for all i=1, …, d and the minimum value, $f(x^*)=0$, occurs at the point x^* = (0,…,0).The function is defined by the following formula:

$$f_{Ra}(x) = 10d + \sum_{i=1}^{d} \left(x_i^2 - 10\cos(2\pi x_i)\right) \quad (5)$$

Styblinski-Tang Function: This is a continuous, multimodal, non-convex function that has been generalized to an n-dimensional domain, with experiments conducted in 2, 5, and 10 dimensions. Locating the minimum of this function is particularly challenging due to its vast search space and complex shape. The function is typically evaluated over the hypercube defined by $x_i \in [-5,5]$ for all i =1,…,d. The minimum occurs at the point x =(2.903534,…,2.903534), with a corresponding value of $f(x)$=39.16599d. The function is defined by the following formula:

$$f_{ST}(x) = \frac{1}{2} + \sum_{i=1}^{d} (x_i^4 - 16x_i^2 + 5x_i) \quad (6)$$

Rosenbrock Function: Also known as the Valley or Banana function, this is a widely used test problem for gradient-based optimization algorithms. The function has been generalized to an n-dimensional domain, with experiments primarily conducted in 2 dimensions. It is unimodal, with the global minimum located in a narrow, parabolic valley. Although this valley is relatively easy to locate, achieving convergence to the minimum is quite challenging. The function is typically evaluated over the hypercube defined by $x_i \in [5,10]$ for all i =1,…,d, with the minimum value, $f(x)$=0, occurring at the point x =(1,…,1).The function is expressed by the following formula:

$$f_{Ro}(x) = \sum_{i=1}^{d-1}\left(100\left(x_{i+1} - x_i^2\right)^2 + (x_i - 1)^2\right) \quad (7)$$

Shubert Function: This is a two-dimensional function characterized by multiple local minima and several global minima. The function is typically evaluated over the square defined by $x_i \in [-10,10]$ for i =1,2. The minimum value of the function is $f(x)$=186.7309. The function can be expressed using the following formula:

$$f_{Sh}(x) = \left(\sum_{i=1}^{5} i\cos((i+1)x_1 + i)\right)\left(\sum_{i=1}^{5} i\cos((i+1)x_2 + i)\right) \quad (8)$$

Some of the selected test functions are designed as minimization problems. In this context, they were transformed into maximization problems by applying negation and adding a constant C. The value of C was determined for each function during the initial experiments. Additionally, the parameters for the Genetic Algorithm were established during these preliminary trials. The following parameter values were used in the experiments:

The genes of individuals are represented by real numbers.

- The crossover probability is set to 0.8.
- The mutation probability is set to 0.15.
- The population size consists of 25 individuals.

Various types of Evolutionary Strategies (ES) were utilized in the experiments, with parameters established during the initial trials. A mutation model was employed that added a random number according to a normal distribution. To adhere to the 1/5 success rule, if no improvement was observed over the course of 10 generations, the mutation size was increased by adjusting the σ\sigmaσ parameter. The number of parents and offspring for the different ES types was determined based on literature recommendations, as shown in Table 1.

Table 1 The number of parents and descendants in different types of Ess.

ES Type	Number of Parents	Number of Descendants
μ+1 ES	20	1
μ, λ, ES	30	200
μ, + λ ES	20	160

The fitness of the best individuals identified by both the Genetic Algorithm (GA) and ES was compared every 50 generations. Based on these comparisons, the top individuals were exchanged between the GA and ES. The algorithm concluded when the best individual reached the predetermined value for the optimized function. Table 2 details the stop criterion value and the constant CCC used for each function. Each algorithm was run 10 times on a standard PC (Intel i3 CPU, 8GB RAM, Windows 10).

Table 2 The value of the algorithm stop criterion and the value of the constant C used for each function.

Function	Stop Criterion	Constant C
f1	1.596	-
f2	0.999	-

f3	2.50048	-
fRa2d	74.999	75
fRa5d	149.999	150
fRa10d	289.5	290
fST2d	656.663	500
fST5d	1641.6	1250
fST10d	2383	2500
fRo	1210121	1210121
fSh	271.795	187

Table 3 presents the average time and the average number of fitness function calls required to reach the target value for the test functions in stage 1 of the experiment. Results from stage 1 indicated that the combination of GA with (1 + 1)-ES demonstrated the highest efficiency. In stage 2, this hybrid algorithm was employed to compare results with the Evolutionary Strategy (ES) and the Standard Genetic Algorithm (SGA), which was adapted for optimizing the test functions. Again, each algorithm was executed 10 times on the same PC configuration. Table 4 shows the average time and the average number of fitness function calls necessary to achieve the predetermined values for the test functions in stage 2. The graph in Figure 2 illustrates the average running time, while Figure 3 depicts the number of fitness function calls for the SGA, ES, and the proposed GA-ES hybrid algorithm.

Table 3 The average running time and the number of fitness function calls needed to reach the predetermined value of the optimized function.

Function	**Algorithm**	**Number of Fitness Calls**	**σ**	**Running Time [s]**	**σ**
f1	GA (1 + 1 ES)	4394	2548	0.000029	0.000014
	GA (μ + 1 ES)	57500	31841	0.000022	0.000011
	GA (μ, λ ES)	197700	156065	0.000194	0.000106
	GA (μ + λ ES)	234696	162480	0.000201	0.000113
f2	GA (1 + 1 ES)	8450	685	0.000048	0.000016
	GA (μ + 1 ES)	205800	26078	0.000069	0.000007
	GA (μ, λ ES)	927700	275635	0.002464	0.000713
	GA (μ + λ ES)	755601	210262	0.001792	0.000483
f3	GA (1 + 1 ES)	7137	2405	0.000072	0.000027
	GA (μ + 1 ES)	840450	548589	0.000158	0.00001
	GA (μ, λ ES)	621400	345817	0.000677	0.000348
	GA (μ + λ ES)	565194	361504	0.000586	0.000336

7. Conclusion

In conclusion, the hybrid multi-evolutionary algorithm developed in this study demonstrates significant promise for addressing complex optimization problems prevalent in data science. By synergizing the exploratory strengths of Genetic Algorithms with the rapid convergence capabilities of Evolutionary Strategies, the proposed GA-ES approach effectively navigates diverse solution landscapes, balancing exploration and exploitation. The experimental results indicate that this hybrid methodology outperforms traditional optimization techniques, particularly in challenging problem domains characterized by multiple local optima. The ability to interchange top-performing individuals between the two algorithms further enhances optimization efficiency, leading to improved solutions in a reduced computational timeframe. Future research could focus on refining the hybrid framework and exploring its applicability to a broader range of real-world data science problems, particularly in areas requiring adaptive optimization strategies. This work underscores the potential of evolutionary algorithms as powerful tools in solving intricate optimization challenges, ultimately contributing to advancements in data-driven decision-making processes.

8. References

1. Yu, Z., Ellahi, R., Nutini, A., Sohail, A., & Sait, S. M. (2021). Modeling and simulations of COVID-19 molecular mechanism induced by cytokines storm during SARS-CoV-2 infection. *Journal of Molecular Liquids, 327*, 114863.
2. Yu, Z., Arif, R., Fahmy, M. A., & Sohail, A. (2021). Self-organizing maps for the parametric analysis of COVID-19 SEIRS delayed model. *Chaos, Solitons & Fractals, 150*, 111202.
3. AbuJarad, M. H., Khan, A. A., Khaleel, M. A., AbuJarad, E. S., AbuJarad, A. H., & Oguntunde, P. E. (2020). Bayesian reliability analysis of Marshall and Olkin model. *Annals of Data Science, 7*(3), 461–489.
4. Sohail, A., Idrees, M., Sajjad, M., Iftikhar, S., & Tunc, S. (2020). Computational framework to explore the impact of environmental stress on epidemics. *Biomedical Engineering (Singapore), 32*, 2050047.
5. Yu, Z., Sohail, A., Nutini, A., & Arif, R. (2020). Delayed modeling approach to forecast the periodic behavior of SARS-CoV-2. *Frontiers in Molecular Biosciences, 7*, 386.
6. Sun, Y., Xue, B., Zhang, M., Yen, G. G., & Lv, J. (2020). Automatically designing CNN architectures using the genetic algorithm for image classification. *IEEE Transactions on Cybernetics, 50*(9), 3854–3864.
7. Lin, H., Yang, C., & Xu, X. (2020). A new optimization model of CCHP system based on genetic algorithm. *Sustainable Cities and Society, 52*, 101811.
8. Drezner, Z., & Drezner, T. D. (2020). Biologically inspired parent selection in genetic algorithms. *Annals of Operations Research, 287*(1), 161–175.
9. Gramaje, A., Thabtah, F., Abdelhamid, N., & Ray, S. K. (2019). Patient discharge classification using machine learning techniques. *Annals of Data Science, 1–13*.
10. Sohail, A. (2019). Inference of biomedical data sets using Bayesian machine learning. *Biomedical Engineering Applications, Basis and Communications, 31*, 1950030.

ISSN: 2249-7196

IJMRR/Sep. 2024/ Volume 14/Issue 7/214-223

Galhenage Gayan Sudesh Suranga Perera/ International Journal of Management Research & Review

11. Iftikhar, M., Sohail, A., & Ahmad, N. (2019). Deterministic and stochastic analysis of dengue spread model. *Biomedical Engineering (Singapore), 31*, 1950008.
12. Chen, N., Xiong, C., Du, W., Wang, C., Lin, X., & Chen, Z. (2019). An improved genetic algorithm coupling a back-propagation neural network model (IGA-BPNN) for water-level predictions. *Water, 11*(9), 1795.