Genetic Algorithm: An Overview And View Through The Scope Of Data Science

Mohamed BENAISSA[*1], Salim KAALDA[2]
[1]Academy of Artificial Intelligence, Faculty of Science Rabat, Morocco.
[2]Artificial Intelligence and Analytics Research Center Sharjah, United Arab Emirates

Article Info :
Accepted: 24/12/2023
Availible online: 11/01/2024

Corresponding Author:
Mohamed BENAISSA, Academy of Artificial Intelligence, Faculty of Science Rabat, Morocco.

ABSTRACT

Of all the types of algorithms in existence, some have the particularity of being inspired by the evolution of species in their natural setting. These are genetic algorithms. Species adapt to their living environment, which may evolve; individuals of each species reproduce, creating new individuals; some undergo DNA modifications; some disappear. In this paper, we explore the world of genetic algorithms, highlighting their principle, their mode of use, their relationship with other machine-learning techniques and their place in the world of decision-making characterized by the emergence of data science and data analytics.

Keywords : *Artificial Intelligence; Genetic Algorithm; Data Science; Machine Learning; Data-driven; Big Data Analytics; Business Analytics; Optimization; Simulation.*

1. INTRODUCTION

A genetic algorithm will reproduce the natural evolutionary model in order to find solutions for an otimization problem whose exact solution is complicated (Figure 1). A genetic algorithm aims to advance a population in order to improve its individuals. With each generation, the focus is on a set of individuals rather than on a single individual. The result is a set of solutions to a problem, rather than a single solution. The solutions identified will often be varied, but of equivalent quality [1].

A genetic algorithm works by evolving a population of individuals over generations, with the aim of improving the overall quality of the population [2]. Unlike some other algorithms, which focus on optimizing a specific individual, a genetic algorithm focuses on a set of individuals at each stage of the evolutionary process [3].

Each generation thus generates a diversity of individuals within the population, each carrying a unique combination of traits or potential solutions to the given problem [4]. Focusing on the population rather than on a single individual encourages the exploration of different avenues and better captures the complexity of the solution space.

The ultimate goal is to obtain a set of diversified and adapted solutions, rather than a single solution [5]. These solutions may differ from one another, but they are all evaluated in terms of their quality. This approach promotes the discovery of robust and effective solutions, offering a holistic and evolutionary perspective to solving complex problems [6-7].

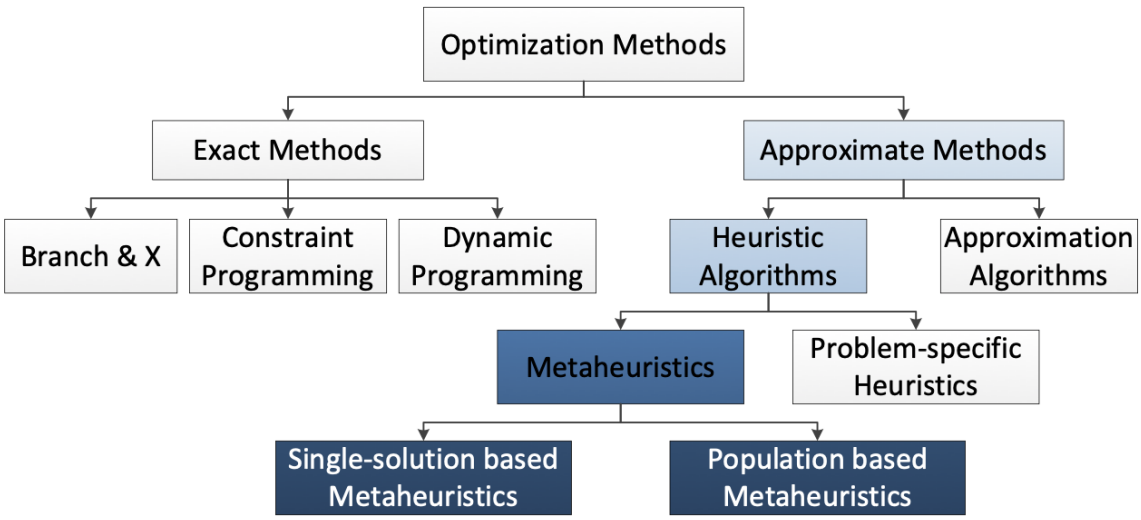

Figure 1: Different optimization methods

In operations research, the Genetic Algorithm is a metaheuristic belonging to the large family of Evolutionary Algorithms, which offer the advantage of providing very high-quality solutions in a reasonable time; the disadvantage is that there is no guarantee that the solution is the global optimum.

2. FROM OPTIMIZATION TO DATA-DRIVEN

The latest technological advances have stimulated the data-driven approach within companies. This involves data-driven business management using digital technologies, with the aim of knowing the state of processes, improving them and providing a more efficient service [8].

Companies are constantly modifying their logistics and production processes to meet changing customer needs. In this sense, data analysis provides detailed information on operational performance and helps identify areas for improvement.

A data-driven company is one whose strategic decision-making is based on objective data rather than experience, intuition or subjective opinions alone [9].

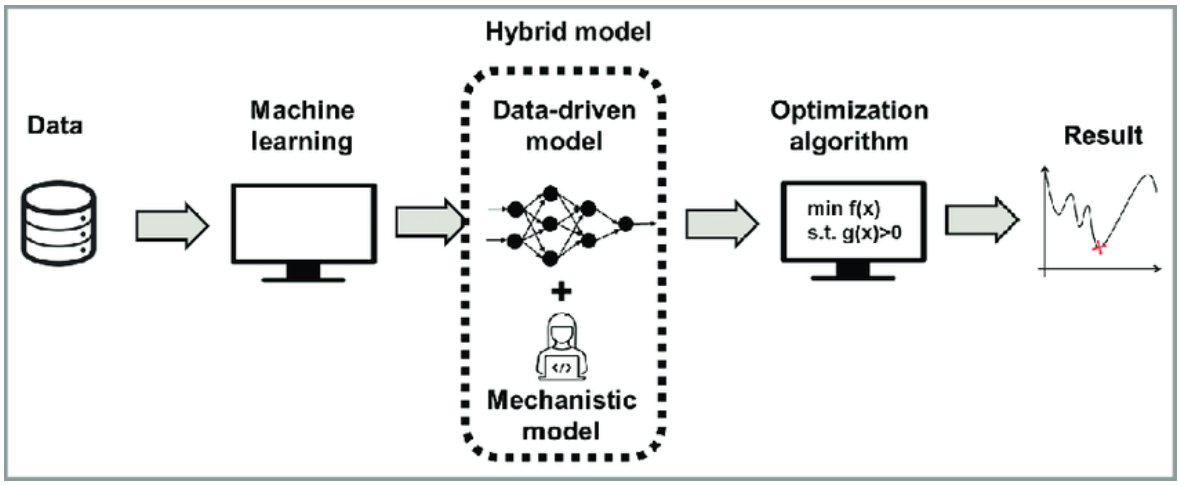

Figure 2: coupling data analysis and optimization [2]

Data-driven is not a new concept. Indeed, companies have always used data analysis to make better decisions. As the British physicist and mathematician William Thomson, better known as Lord Kelvin, once said, "you can only improve what you measure". Data acts as a business driver, since by measuring processes, we obtain a detailed mapping of activity, enabling us to introduce improvements [10].

Over the last few decades, the emergence of new technologies such as Big Data and artificial intelligence has enabled companies to gain greater precision in analyzing their operations. Data-driven organizations are choosing to digitize their business in order to make the most of data. Collecting, analyzing and sharing information contributes to cooperation between different departments [11]. According to consulting firm Gartner, by 2025, 95% of decisions based on data today will be at least partially automated.

One of the benefits of the data-driven approach is the increased transparency and visibility of processes. Using key performance indicators or KPIs, companies can accurately monitor logistics and production processes, anticipate certain situations and take proactive measures to improve efficiency. "Preventive maintenance in particular plays a decisive role in reducing costs and improving business performance. Using heterogeneous data sources, it is possible to detect equipment anomalies (diagnosis), predict potential failures (prognosis) and support anticipated decisions (proactive decision-making)", say the authors of the study.

The intersection between genetic algorithms and machine learning (Figure 3) lies in the use of genetic algorithms as an optimization technique within the overall machine learning process. Genetic algorithms can be used to efficiently search for the best combinations of hyperparameters in a machine learning model. This makes it possible to improve model performance by adjusting parameters such as learning speed, neural network depth, or other variables related to the learning algorithm. Some research is exploring the use of genetic algorithms to directly evolve the architecture of machine learning models. This can include mutating and recombining layers of neural networks to create more efficient architectures. By combining genetic algorithms with machine learning, it is possible to create models that adapt dynamically to changes in input data, enabling greater resilience in the face of changing conditions.

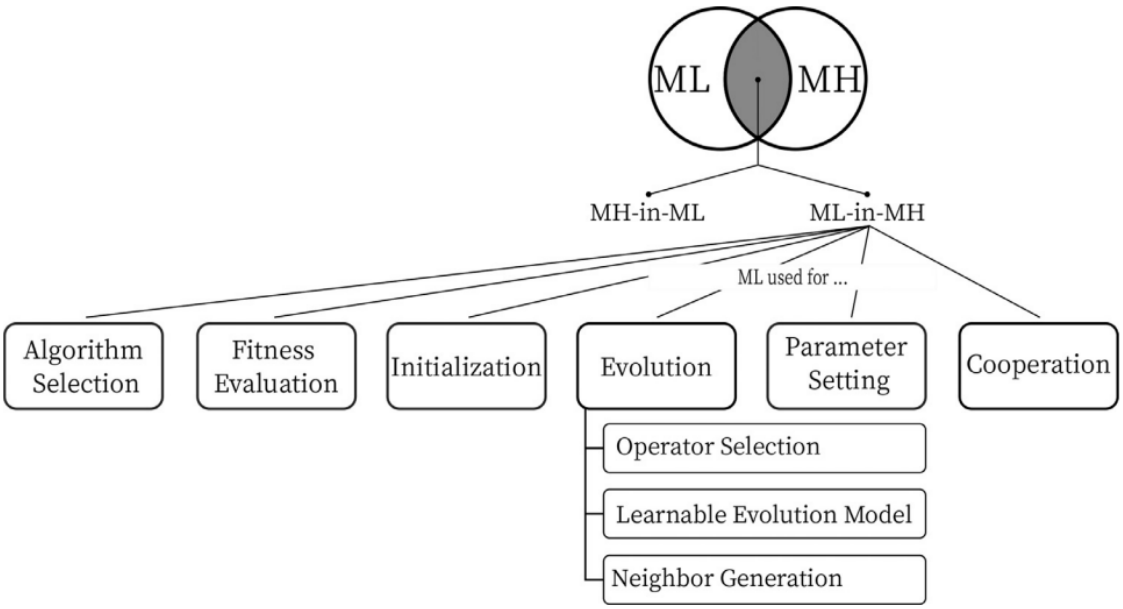

Figure 3: Taxonomy on the use of ML in MHs (ML-in-MH) [4]

3. FRAMEWORK FOR COUPLING GENITIC ALAGORITHM WITH DATA ANALYTICS

The framework shown in the following figure (Figure 4) illustrates the relationship between the two concepts of data science and optimization. It consists of:

- Data collection module: the data collection module refers to the methods used to collect different forms of data. structured data such as RDBMS or unstructured data such as log files, videos..ect. Standard data collection techniques require the use of real-time or batch collection tools.
- Data storage module : it brings together a body of information that is organized for easy access, management and updating by its users. Data is organized in rows, columns and tables, and indexed for easy retrieval. Companies use databases to store, manage and retrieve information. In addition, the module offers the possibility of storing unstructured data in a Data Lake. It offers the possibility of storing a very large quantity of raw or highly refined data for a given period of time. The Data Lake is one of a number of backup solutions that facilitate the cohabitation of different data structures and schemas. In other words, this storage method promotes cohabitation between copies of source system data, raw data and transformed data.
- Data analytics module : a module for examining raw structured or unstructured data, with the aim of drawing conclusions from this information. The Data Analytics module uses visualization and statistical techniques to help companies and organizations make better decisions.
- Optimization module: an optimization module is a software component that offers specific functionalities designed to improve the performance, efficiency or quality of a system or application. Optimization in this context concerns the application of operational research techniques to find an optimized solution from the data predicted by the analytics module.

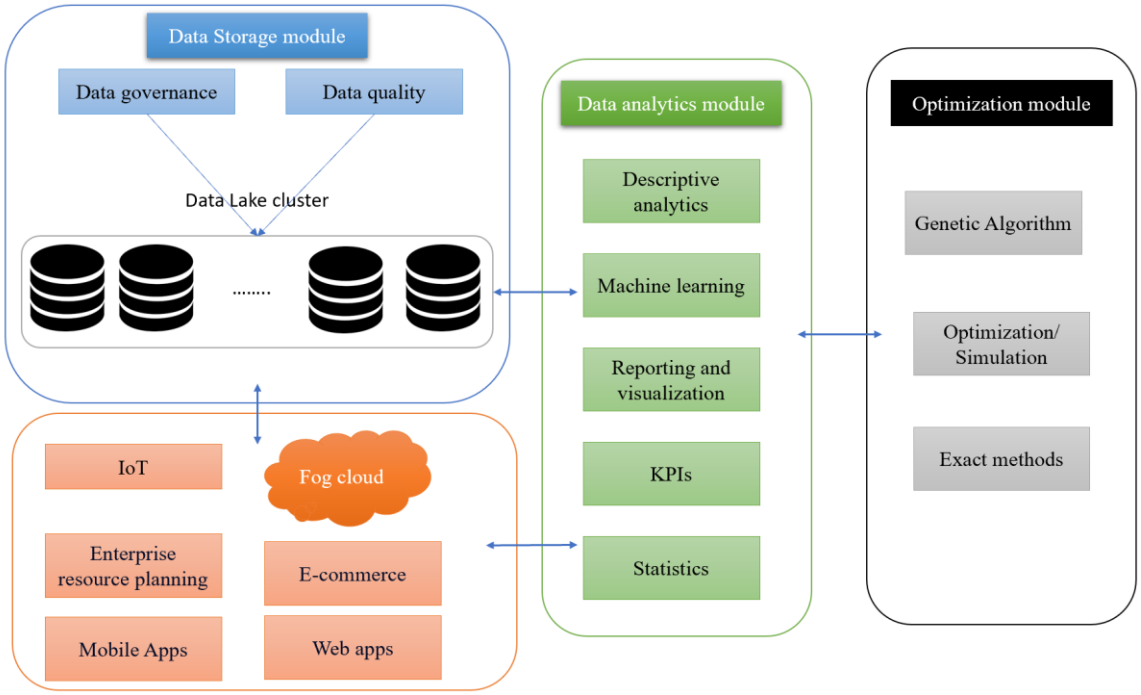

Figure 4: framework coupling genitic alagorithm & data analytics

4. CONCLUSION

In conclusion, genetic algorithms stand out among the diverse array of algorithms by drawing inspiration from the evolutionary processes observed in natural species. This unique approach mirrors the adaptability seen in biological systems, where species evolve in response to changes in their environment. Through the reproduction of individuals, genetic algorithms generate new solutions, with some undergoing DNA modifications while others may fade away.

Throughout this paper, we delved into the realm of genetic algorithms, shedding light on their fundamental principles, practical applications, and their interplay with other machine-learning techniques. As decision-making processes evolve in the era marked by the rise of data science and data analytics, genetic algorithms find their place as powerful tools capable of exploring solution spaces, optimizing parameters, and contributing to the adaptive and intelligent nature of modern systems. Embracing the evolutionary concepts that underlie genetic algorithms can open new avenues for addressing complex problems and harnessing the potential of artificial intelligence in the ever-evolving landscape of computational methodologies.

CONFLICTS OF INTEREST

All authors declare that they have no conflicts of interest.

REFERENCES

[1] Claudia Cavallaro, Vincenzo Cutello, Mario Pavone, Francesco Zito, Machine Learning and Genetic Algorithms: A case study on image reconstruction, Knowledge-Based Systems, Volume 284, 2024, 111194, ISSN 0950-7051, https://doi.org/10.1016/j.knosys.2023.111194.

[2] Schweidtmann, A.M.; Esche, E.; Fischer, A.; Kloft, M.; Repke, J.U.; Sager, S.; Mitsos, A. Machine Learning in Chemical Engineering: A Perspective. Chem. Ing. Tech. 2021, 93, 2029–2039.

[3] Maia, M.R.H., Reula, M., Parreño-Torres, C. et al. Metaheuristic techniques for the capacitated facility location problem with customer incompatibilities. Soft Comput 27, 4685–4698 (2023). https://doi.org/10.1007/s00500-022-07600-z

[4] Maryam Karimi-Mamaghan, Mehrdad Mohammadi, Patrick Meyer, Amir Mohammad Karimi-Mamaghan, El-Ghazali Talbi, Machine learning at the service of meta-heuristics for solving combinatorial optimization problems: A state-of-the-art, European Journal of Operational Research, Volume 296, Issue 2, 2022, Pages 393-422, ISSN 0377-2217, https://doi.org/10.1016/j.ejor.2021.04.032.

[5] Vasileios A. Tatsis, Konstantinos E. Parsopoulos, Dynamic parameter adaptation in metaheuristics using gradient approximation and line search, Applied Soft Computing, Volume 74, 2019, Pages 368-384, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2018.09.034.

[6] Abdullah Konak, David W. Coit, Alice E. Smith, Multi-objective optimization using genetic algorithms: A tutorial, Reliability Engineering & System Safety, Volume 91, Issue 9, 2006, Pages 992-1007, ISSN 0951-8320, https://doi.org/10.1016/j.ress.2005.11.018.

[7] Xin-She Yang, Chapter 5 - Genetic Algorithms, Editor(s): Xin-She Yang, Nature-Inspired Optimization Algorithms, Elsevier, 2014, Pages 77-87, ISBN 9780124167438, https://doi.org/10.1016/B978-0-12-416743-8.00005-1.

[8] OLIVA, Diego, HOUSSEIN, Essam H., et HINOJOSA, Salvador (ed.). Metaheuristics in machine learning: theory and applications. Berlin : Springer, 2021.

[9] Samuel Fosso Wamba, Maciel M. Queiroz, Laura Trinchera, The role of artificial intelligence-enabled dynamic capability on environmental performance: The mediation effect of a data-driven culture in France and the USA, International Journal of Production Economics, Volume 268, 2024, 109131, ISSN 0925-5273, https://doi.org/10.1016/j.ijpe.2023.109131.

[10] AWAN, Usama, BRAATHEN, Petter, et HANNOLA, Lea. When and how the implementation of green human resource management and data-driven culture to improve the firm sustainable environmental development?. Sustainable Development, 2023.

[11] Smail Benzidia, Naouel Makaoui, Omar Bentahar, The impact of big data analytics and artificial intelligence on green supply chain process integration and hospital environmental performance, Technological Forecasting and Social Change, Volume 165, 2021, 120557, ISSN 0040-1625, https://doi.org/10.1016/j.techfore.2020.120557.