2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA) | 979-8-3503-4503-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/DSAA60987.2023.10302494

Hybrid Approaches to Optimization and Machine Learning Methods

1st Beatriz Flamia Azevedo
Research Centre in Digitalization and Intelligent Robotics (CeDRI/SusTEC), Instituto Politécnico de Bragança, Bragança, Portugal
ALGORITMI Research Centre/LASI, University of Minho, Braga, Portugal
beatrizflamia@ipb.pt

2nd Ana Maria A. C. Rocha
ALGORITMI Research Centre/LASI
University of Minho
Braga, Portugal
arocha@dps.uminho.pt

3rd Ana I. Pereira
Research Centre in Digitalization and Intelligent Robotics (CeDRI/SusTEC)
Instituto Politécnico de Bragança
Bragança, Portugal
apereira@ipb.pt

***Abstract*—This paper conducts a comprehensive literature review concerning hybrid techniques that combine optimization and machine learning approaches for clustering and classification problems. The aim is to identify the potential benefits of integrating these methods to address challenges in both fields. The paper outlines optimization and machine learning methods and provides a quantitative overview of publications since 1970. Additionally, it offers a detailed review of recent advancements in the last three years. The study includes a SWOT analysis of the top ten most cited algorithms from the collected database, examining their strengths and weaknesses as well as uncovering opportunities and threats explored through hybrid approaches. Through this research, the study highlights significant findings in the realm of hybrid methods for clustering and classification, showcasing how such integrations can enhance the shortcomings of individual techniques.**

***Index Terms*—machine learning, optimization, hybrid methods, literature review, clustering, classification.**

I. Introduction

In addressing the growing complexity of real-world challenges and the need for efficient solutions for large datasets, the demand for advanced models and algorithms has risen, both in optimization and machine learning. A hybrid algorithm can combine optimization and machine learning techniques is an effective strategy that uses the advantages of both methodologies to provide a powerful framework for tackling complex problems. This fusion improves decision-making by blending optimization into machine learning and vice versa. Thereby, a hybrid algorithm can leverage optimization capabilities to guide the learning process and enhance the accuracy and efficiency of decision-making. This integration empowers the algorithm to use mathematical optimization and learning, resulting in better decision-making [2]–[4].

The authors are grateful to the Foundation for Science and Technology (FCT, Portugal) for financial support through national funds FCT/MCTES (PIDDAC) to CeDRI (UIDB/05757/2020 and UIDP/05757/2020), Algoritmi (UIDB/00319/2020) and SusTEC (LA/P/0007/2021). Beatriz Flamia Azevedo is supported by FCT Grant Reference SFRH/BD/07427/2021.

This paper describes and explores the main characteristics of numerical optimization and machine learning methods. Furthermore, a deep and systematic literature review looks at how these methods have evolved and analyze how they can be combined to overcome challenges. The goal is to find ways to make them work better together, like using ideas from one method to improve the other. Although the work presents the different types of machine learning in detail, the literature review will be restricted to algorithms that perform the classification or clustering task.

In this way, this paper makes a significant contribution by systematically identifying and analyzing the existing knowledge on hybrid algorithms that combine optimization and machine learning. It finds gaps in research and suggests future directions. Additionally, the paper presents a comprehensive SWOT analysis of the top ten most cited algorithms in the collected database. Consequently, the paper offers a thorough understanding of the characteristics of these algorithms, serving as a valuable source of inspiration for future research endeavors.

II. Methodology and Data Base

The methodology consists of an extensive overview and bibliometric analysis of the literature on hybrid approaches involving optimization and machine learning algorithms. Thus, a theoretical foundation on optimization and machine learning techniques were presented. Initially, a historical survey of works published since 1970 was carried out, involving the main themes contemplated in this work: Optimization, Machine Learning, Swarm, and Evolutionary algorithms, in order to analyze the evolution of the theme throughout the years. To generate the information need it, a logic search was defined and applied to three databases (Scopus, IEEE, and WoS) together with some constraints of year, language, and publisher type. So, through an explanatory analysis, the following logic search was defined: "(Machine learning) AND (Optimization)

AND (Swarm OR Evolutionary) AND (Classification OR Clustering) AND NOT (Reinforcement OR Neural Networks OR Ensemble methods OR Regression OR Game theory OR Robotics OR Deep learning)". Thereby, 1007 papers published between 2019 and 2021, provided by three databases (479 from Scopus, 200 from IEEE, and 450 WoS). They were analyzed in a systematic and bibliometric way, which resulted in an extensive range of works. Finally, a SWOT analysis of the most frequent algorithms in the data set was performed to identify the strengths, weaknesses, opportunities, and threats of each one and point out some works that explore such aspects.

III. Summary

The systematic review has made valuable contributions to several hybrid approaches encompassing optimization and machine learning techniques. Each algorithm presented in this context demonstrates its own set of advantages and drawbacks, demanding careful evaluation to select the most suitable strategies for addressing specific problems. Among the many aspects explored within these algorithms, two key elements are particularly noteworthy: parameter estimation and search technique enhancement [3]–[5], both crucial for avoiding local optima. Relying on empirically determined initial parameters is considered a drawback due to its subjectivity and potential bias, which can compromise algorithm performance. Similarly, becoming trapped in local optima is highly unfavorable for global optimization algorithms, leading to erroneous solutions. Conversely, the prospects and challenges in this field are closely linked to the identified strengths and weaknesses. Exploring these aspects facilitates the mitigation of weaknesses by harnessing the potential of other methods, such as hybrid approaches. Moreover, the challenges posed by one algorithm can be effectively tackled by the strengths of alternative algorithms, as highlighted in the No-Free-Lunch theorem [1].

The performance of an algorithm depends on several factors, as the solution quality and consumed budget are the most significant measures in the algorithm performance assessment. There is no universal and perfect method and/or algorithm to solve all optimization problems. Based on this, it is a hard task to define the best strategy (or algorithm) without considering the problem and the data information, leaving the option of analyzing the characteristics of all and finding ways to choose the most appropriate one.

Throughout this work, several research were explored in order to explore the methods or techniques already developed that combine optimization and machine learning in order to improve algorithms performance. Among these works, the strong presence of strategies that used the PSO algorithm stands out, due to its speed and robustness, and exploration capabilities, to strengthen both optimization and machine learning algorithms; thus, it is the most mentioned algorithm in the performed search. Due to the large number of works found and their peculiarities, it is a hard task to list all combinations of algorithms found. However, the SWOT analysis could provide more details about the Strengths, Weaknesses, Opportunities, and Threats of the most cited algorithms.

In general, using hybrid tools allows some of the difficulties encountered by the methods to be mitigated or eliminated. Therefore, the results that have been obtained with hybridization are very promising. Most of the time, hybrid algorithms outperform pure algorithms in several aspects, such as speed, exploration and/or exploitation abilities, accuracy, among other aspects, without requiring more computational resources than pure methods. Furthermore, the estimation of parameters, one of the most mentioned obstacles, which is done empirically in pure methods, is now based on knowledge of the data, which provides a result that is often superior and more appropriate for the learning task, free from any possible bias.

Given the significance of hybrid methods in addressing practical challenges, numerous unexplored opportunities exist to leverage the synergy between optimization and machine learning. Looking ahead to future advancements and literature exploration, it would be compelling to explore variations of established hybrid approaches. This comparative assessment of hybrid technique performance and adaptability across diverse problem domains could unveil insightful findings for further advancement. To conclude, due to the stringent search criteria applied in our review, it's possible that some crucial contributions may have been overlooked within the extensive literature. Therefore, in forthcoming research, it is advisable to expand the scope of the literature search to individually study some method categories, such as supervised, unsupervised, reinforcement, evolutionary algorithms, and swarm algorithms. Adopting this approach could explore more deeply the literature review, leading to a more comprehensive discussion of the merits and limitations of cutting-edge techniques.

References

[1] Wolpert, D. & Macready, W. "No Free Lunch Theorems for Optimization." IEEE Trans. On Evolutionary Computation. **1** pp. 67-82, 1997.

[2] Goldberg, D. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley Professional, 1989.

[3] Karaboga, D., Akay, B. & Karaboga, N. "A Survey on the Studies Employing Machine Learning (ML) For Enhancing Artificial Bee Colony (ABC) Optimization Algorithm." Cogent Engineering. **7** pp. 1855741, 2020.

[4] Ezugwu, A., Shukla, A., Agbaje, M., Oyelade & Agushaka, J. "Automatic Clustering Algorithms: A Systematic Review and Bibliometric Analysis of Relevant Literature". Neural Computing And Applications. **33** pp. 6247-6306, 2021.

[5] Tharwat, A. & Gabel, T. "Parameters Optimization of Support Vector Machines for Imbalanced Data Using Social Ski Driver Algorithm." Neural Comput. Appl. **32** pp. 6925-6938, 2020.