Faculty of Sciences

Department of Computer Science and Information Systems

Rhino Optimization Algorithm (ROA): A MetaHeuristic Optimization Algorithm for Hyperparameter Optimization

by

SIMISANI NDABA

Student ID number: 24020000

BIS (Business) (UoB), MSc (Computer Information Systems) (UoB)

A Concept Paper Submitted to the Faculty of Sciences in Partial Fulfilment of the Requirements for the Award of the Degree of Doctor of Philosophy in Information Systems of BIUST

Supervisor(s): Prof, Rajalakshmi Selvaraj

Department of Computer Science and Information Systems

Faculty of science, BIUST

E-mail Address: selvarajr@biust.ac.bw

Dr. Hlomani Hlomani

Department of Computer Science and Information Systems

Faculty of science, BIUST

E-mail Address: hlomanihb@biust.ac.bw

September, 2024

1. SECTION 1. INTRODUCTION

A learning algorithm's hyperparameters are the parameters whose values regulate the learning process and determine the models' final parameters. Finding the best settings for hyperparameters to get good results from data as quickly as possible is the purpose of hyperparameter optimization (HPO) (Li et al ,2021).

Hyperparameter optimization (HPO) is an important subfield of machine learning that focuses on tuning the hyperparameters of a chosen algorithm to achieve peak performance (Kadr et al., 2024). Hyperparameter optimization and fine-tuning methods are crucial components in the optimisation of machine learning models, essential for ensuring their adaptability and effectiveness across diverse datasets and tasks (Snoek et al.,2012). An algorithm that attempts this task is called a hyperparameter optimization (HPO) procedure (Claesen and De Moor., 2015; Feurer and Hutter., 2019).

In recent years, nature-inspired/Metaheuristic/Bio-optimization algorithm (BoA) approaches have gained popularity due to local and global convergence and have been applied to many optimization problems. Bio-optimization algorithms (BoAs) follow a two-step process, consisting of the exploration and exploitation stages. Due to their higher convergence speed, few parameters, and simple implementation, these BoAs are widely used in solving parameter optimization, fault prediction, bioinformatics, image processing, and many other real-time engineering problems (Trojovská et al., 2022).

Background

Yu and Zhu (2020) explain the history of HPO dates back to the early 1990s as stated by Ripley (1993) and King et al (1995), and the method is widely applied for neural networks with the increasing use of machine learning. HPO can be viewed as the final step of model design and the first step of training a neural network. Considering the influence of hyper-parameters on training accuracy and speed, they must carefully be configured with experience before the training process begins (Rodriguez, 2018).

Conceptually, HPOs purposes are threefold, Feurer and Hutter (2019) say that to reduce the costly menial work of artificial intelligence (AI) experts and lower the threshold of research and development; to improve the accuracy and efficiency of neural network training (Melis et al., 2017);

and to make the choice of hyper-parameter set more convincing and the training results more reproducible (Bergstra et al., 2013).

Trojovská et al (2022) say global optimization problems in real applied optimization problems have high dimensions and complexity because they involve many multiple decision variables and many complex nonlinear relationships. The complex and non-convex nature of the problems and their unknown search space in real-world applications make analytical methods almost unusable (Agrawal et al., 2021). The weakness of analytical mathematical methods in solving optimization problems has led to the creation of a special type of intelligent search algorithms called meta-heuristic algorithms. Meta-heuristic algorithms are stochastic methods that, inspired by nature and its mechanisms, try to send their initial population to the global optimum and provide appropriate solutions close to the global optimum in a reasonable time (Agrawal et al., 2021).

Trojovská et al (2022) further explain optimization problem-solving methods fall into two groups: deterministic methods and stochastic methods. Deterministic methods in both gradient-based and non-gradient-based groups have good performance in dealing with linear, convex, continuous, and differentiable optimization problems (Agrawal et al., 2021). On the other hand, many real-world optimization problems have properties such as nonlinear, non-convex, non-differentiable objective function, discrete search space, and high dimensions. These features have led to the failure of deterministic methods in handling such optimization applications and thus led to the introduction of another approach called stochastic methods (Zhang et al., 2021).

A study from Wu et al (2019) explains that Stochastic methods in the optimization process use random search in the problem-solving space based on search agents, random operators, and trial and error processes. Metaheuristic algorithms are one of the most efficient stochastic methods that, based on collective intelligence, provide effective performance in optimizing and providing solutions to optimization problems. Based on the fact that these solutions may not be the same as the global optima of optimization problems, Zhang et al (2021) describe the solutions obtained from metaheuristic algorithms as quasi-optimal.

Studies by Dehghani et al, 2021 and Trojovská and Dehghani (2022) have found that the desire to achieve better quasi-optimal solutions has been the primary motivation for researchers to design numerous metaheuristic algorithms. As a result, numerous metaheuristic algorithms have been developed based on the simulation of evolution-based processes in nature, physical phenomena, biological sciences, animal behavior, and other living organisms have been used.

Some of the most well-known optimization techniques are Differential evolution (DE) by Storn and Price (1997), Harmony search (HS) by Geem et al., 2001), Ant colony optimization (ACO) by Dorigo and Di Caro(1999), Firefly algorithm (FA) by Yang (2010), Cuckoo search (CS) Yang and Deb (2009), Gravitational search algorithm (GSA) by Rashedi et al (2009), Grey wolf optimizer (GWO) by Mirjalili et al (2014). To some extent, the algorithms mentioned above are inspired by some, such as the social behavior of animal groups (foraging, migration, courtship), the evolution of nature, human and social behavior.

Problem Statement

Traditional HPO approaches such as Grid Search, Random Search, and Bayesian Optimization have been widely used in hyperparameter optimisation. However, as datasets grow and models increase in complexity, these approaches often require a significant amount of time and resources for HPO (Khadka et al., 2024). The hyperparameter problem stated by Strauss (2007) is to determine which hyper-parameters to tune and their search space, adjust them from coarse to fine, evaluate the performance of the model with different parameter sets, and determine the optimal combination.

Despite the development of numerous BoAs, each algorithm possesses its own set of limitations. Based on the No Free Lunch (NFL) theorem described by Hussien et al (2020), no single BoA can perform best in addressing all optimization problems. As a result, researchers are interested in exploring and proposing effective optimization methods to tackle a broader range of optimization problems (Das et al, 2024). Therefore, Dehghani et al (2021) find that this issue is resolved by improving existing methods or introducing newer optimization algorithms. An important issue in improving the capability of optimization algorithms is to increase the exploration power to global search the problem-solving space and to increase the exploitation power to local search the optimal area discovered, while a proper balance must be struck between these two indicators.

Contribution

This paper's novelty and scientific contribution are to design a new metaheuristic swarm-based algorithm called Rhino Optimization Algorithm (ROA). Rhinos are land animals whose social behaviour, forging and running away from predators strategy represents an optimization process. ROA's fundamental inspiration is to model the social behavior of herds of rhinos in the wild. To the best of the authors' knowledge of the literature, no optimization algorithm has been developed based on the Rhino behavior. This research gap is the motivation to develop a new optimization algorithm by mathematically modeling the Rhino's behaviour.

Justification of the study

Various Swarm optimisers have been proposed recently. Most of these approaches are inspired by foraging, mating, hunting and searching behaviors of animals in nature. In the scope of our knowledge, there is no swarm optimiser method in the literature inspired by the social behaviors of Rhinos.

General Objective

1. Inspired by the social behaviors of the Rhino swarm which summarized observations in daily life, a novel group-based swarm intelligence algorithm is proposed, named Rhino Optimisation Algorithm (ROA).
2. Two rules are modeled from the finding food and foraging of the duck, which corresponds to the exploration and exploitation phases of the proposed ROA, respectively.
3. Compare the proposed approach with existing metaheuristic swarm HPOs.

Specific Objectives

1. Create a novel algorithm which can be used as part of the existing optimisers.
2. Combine different hyper-parameter optimisation algorithms (optimisers) by combining their functionality, ultimately creating a new algorithm that can contribute to the highest model performance.
3. Formulate a mathematical simulation of a Rhinos' life for the optimisation process.
4. Use the proposed ROA on deep learning algorithms namely, Convolutional Neural Network (CNN), Recurrent neural networks (RNNs) and Long Short-Term Memory Networks (LSTMs).

Research Questions

1. Based on the No Free Lunch (NFL) Theorem, is there still a need to design new optimization algorithms?
2. Can the ROA achieve better quasi-optimal solutions?
3. Can the nature inspired Rhino behavioural pattern be a solution for hyperparameter optimisation?

5. SECTION 2. LITERATURE REVIEW

This literature review examines various journal articles, providing a comprehensive overview of the current status of research, identifying existing research gaps, and justifying the research methods in hyper-parameter optimisation. The related work is sectioned into the types of work hyperameter optimisation has been studied. The sections include, Comparitice HPO, proposed HPOs, Combined HPOs, combined HPO and ML and improved HPOs

Creating new HPOs
Mallik et al 2024 worked on PriorBand, an HPO algorithm for DL researchers. They added the incumbent-based sampling to existing uniform sampling, and prior-based sampling that are in the Hyperband algothim

6. SECTION 3. THEORETICAL FUNDAMENTALS / METHODOLOGY

Theoretical Framework

For the experiment to work systematically, I propose to use an experimental research method. I plan on first following the following procedure.
The focus is to conduct experiments to test hypotheses or explore model behaviour.
Steps:
1. Hypothesis Formulation: Formulate a clear hypothesis or research question.
2. Experimental Design: Design experiments to test my hypothesis, including choosing datasets and metrics.
3. Implementation: Write code to conduct the experiments.
4. Execution: Run experiments and collect data.
5. Analysis: Analyze the results statistically to draw conclusions.
6. Reproducibility: Ensure my experiments can be reproduced by others.

Performance Metrics
Kareem et al (2024) explain and show the equation for,
Root Mean Square Error (RMSE)
RMSE is the average squared deviation between the estimated and measured outputs. It is utilised for assessing the nonlinear error; this is an excellent measure of forecast accuracy [78,138].
The Equation for determining RMSE is exposed in Eq. (1):
RMSE =

N
i
(Xi − Yi)2
N
(1)
where Yi is: the forecast value, Xi is the actual value, X : mean of the actual value, Y : the mean of

the predicted value, N: the total number, i: counter.

Mean Absolute Error (MAE)

MAE estimates the mean of error magnitudes without regard to their direction. In other words, this is the mean absolute deviation between the predicted and actual values [142].

MAE = 1
N
N
i=1
|Xi − Yi| (2)

REFERENCES

Agrawal, P., Abutarboush, H. F., Ganesh, T., & Mohamed, A. W. (2021). Metaheuristic algorithms on feature selection: A survey of one decade of research (2009-2019). Ieee Access, 9, 26766-26791. https://ieeexplore.ieee.org/abstract/document/9344597.

Alibrahim, H., & Ludwig, S. A. (2021, June). Hyperparameter optimization: Comparing genetic algorithm against grid search and bayesian optimization. In 2021 IEEE Congress on Evolutionary Computation (CEC) (pp. 1551-1559). IEEE.

Afape, J. O., Willoughby, A. A., Sanyaolu, M. E., Obiyemi, O. O., Moloi, K., Jooda, J. O., & Dairo, O. F. (2024). Improving millimetre-wave path loss estimation using automated hyperparameter-tuned stacking ensemble regression machine learning. Results in Engineering, 22, 102289.

Bergstra, J., Yamins, D., & Cox, D. (2013, February). Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In International conference on machine learning (pp. 115-123). PMLR.https://proceedings.mlr.press/v28/bergstra13.html.

Cavalcanti, J. H., Kovács, T., & Kő, A. (2022). Production system efficiency optimization using sensor data, machine learning-based simulation and genetic algorithms. Procedia CIRP, 107, 528-533.

Claesen, M., & De Moor, B. (2015). Hyperparameter search in machine learning. arXiv preprint arXiv:1502.02127.

Coelho, R. C., Alves, A. F. C., & Pires, F. A. (2024). Efficient constitutive parameter identification through optimisation-based techniques: A Comparative analysis and novel composite Bayesian optimisation strategy. Computer Methods in Applied Mechanics and Engineering, 427, 117039. https://doi.org/10.1016/j.cma.2024.117039

Das, M., Mohan, B. R., Guddeti, R. M. R., & Prasad, N. (2024). Hybrid Bio-Optimized Algorithms for Hyperparameter Tuning in Machine Learning Models: A Software Defect Prediction Case Study. Mathematics, 12(16), 2521.https://www.mdpi.com/2227-7390/12/16/2521

Dehghani, M., Hubálovský, Š., & Trojovský, P. (2021). Northern goshawk optimization: a new swarm-based algorithm for solving optimization problems. Ieee Access, 9, 162059-162080.https://ieeexplore.ieee.org/abstract/document/9638618.

Dorigo, M., & Di Caro, G. (1999, July). Ant colony optimization: a new meta-heuristic. In Proceedings of the 1999 congress on evolutionary computation-CEC99 (Cat. No. 99TH8406) (Vol. 2, pp. 1470-1477). IEEE.https://ieeexplore.ieee.org/abstract/document/782657.

Feurer, M., & Hutter, F. (2019). Hyperparameter optimization. *Automated machine learning: Methods, systems, challenges*, 3-33.

Geem, Z. W., Kim, J. H., & Loganathan, G. V. (2001). A new heuristic optimization algorithm: harmony search. simulation, 76(2), 60-68.https://doi.org/10.1177/003754970107600201.

Hanifi, S., Cammarono, A., & Zare-Behtash, H. (2024). Advanced hyperparameter optimization of deep learning models for wind power prediction. Renewable Energy, 221, 119700.

Hussien, A. G., Amin, M., Wang, M., Liang, G., Alsanad, A., Gumaei, A., & Chen, H. (2020). Crow search algorithm: theory, recent advances, and applications. IEEE Access, 8, 173548-173565.y

Kadra, A., Janowski, M., Wistuba, M., & Grabocka, J. (2024). Scaling laws for hyperparameter optimization. Advances in Neural Information Processing Systems, 36.

Khadka, K., Chandrasekaran, J., Lei, Y., Kacker, R. N., & Kuhn, D. R. (2024, April). A Combinatorial Approach to Hyperparameter Optimization. In Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering-Software Engineering for AI (pp. 140-149).

King, R. D., Feng, C., & Sutherland, A. (1995). Statlog: comparison of classification algorithms on large real-world problems. Applied Artificial Intelligence an International Journal, 9(3), 289-333.https://doi.org/10.1080/08839519508945477

Lin, J. A., Padhy, S., Mlodozeniec, B., Antorán, J., & Hernández-Lobato, J. M. (2024). Improving Linear System Solvers for Hyperparameter Optimisation in Iterative Gaussian Processes. arXiv preprint arXiv:2405.18457. https://doi.org/10.48550/arXiv.2405.18457.

Li, Y., Zhang, Y., & Cai, Y. (2021). A new hyper-parameter optimization method for power load forecast based on recurrent neural networks. Algorithms, 14(6), 163. https://doi.org/10.3390/a14060163.

Martin J. Strauss. Typical steps for solving optimization problems. Electrical Engineering and Computer Science, University of Michigan, 2007.

Melis, G., Dyer, C., & Blunsom, P. (2017). On the state of the art of evaluation in neural language models. arXiv preprint arXiv:1707.05589. https://doi.org/10.48550/arXiv.1707.05589.

Mirjalili S, Gandomi AH, Mirjalili SZ, Saremi S, Faris H, Mirjalili SM (2017) Salp Swarm algorithm: a bio-inspired optimizer for engineering design problems. Adv Eng Softw 114:163–191

Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. Advances in engineering software, 69, 46-61.https://doi.org/10.1016/j.advengsoft.2013.12.007.

Nguyen, D. H., Le, X. H., Anh, D. T., Kim, S. H., Bae, D. H. (2022). Hourly streamflow forecasting using a Bayesian additive regression tree model hybridized with a genetic algorithm. Journal of Hydrology, 606(3–4), 127445. https://doi.org/10.1016/j.jhydrol.2022.127445.

Rashedi, E., Nezamabadi-Pour, H., & Saryazdi, S. (2009). GSA: a gravitational search algorithm. Information sciences, 179(13), 2232-2248.https://doi.org/10.1016/j.ins.2009.03.004.

Ripley, B. D. (1993). Statistical aspects of neural networks. *Natworks and Chaos-Statistical and Probabilistic Aspects*, 40-123.

Rodriguez, J. (2018). Understanding hyperparameters optimization in deep learning models: concepts and tools. *Linkedin Pulse*.

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. In Advances in neural information processing systems (pp. 2951-2959).

Storn, R., & Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. Journal of global optimization, 11, 341-359., https://link.springer.com/article/10.1023/a:1008202821328.

Tian, J., Zhang, X., Zheng, S., Liu, Z., & Zhan, C. (2024). Synergising an Advanced Optimisation Technique with Deep Learning: A Novel Method in Fault Warning Systems. Mathematics, 12(9), 1301.

Trojovská, E., Dehghani, M., & Trojovský, P. (2022). Zebra optimization algorithm: A new bio-inspired optimization algorithm for solving optimization algorithm. Ieee Access, 10, 49445-49473.https://ieeexplore.ieee.org/abstract/document/9768820

Trojovská, E., & Dehghani, M. (2022). Clouded leopard optimization: a new nature-inspired optimization algorithm. IEEE Access, 10, 102876-102906.

Yang, X. S., & Deb, S. (2009, December). Cuckoo search via Lévy flights. In 2009 World congress on nature & biologically inspired computing (NaBIC) (pp. 210-214). 10.1109/NABIC.2009.5393690.

Yang, X. S. (2010). Nature-inspired metaheuristic algorithms. Luniver press.

Yu, T., & Zhu, H. (2020). Hyper-parameter optimization: A review of algorithms and applications. arXiv preprint arXiv:2003.05689. https://doi.org/10.48550/arXiv.2003.05689

X. Wu, S. Zhang, W. Xiao, and Y. Yin, ``The exploration/exploitation tradeoff in whale optimization algorithm," IEEE Access, vol. 7, pp. 125919125928, 2019.https://ieeexplore.ieee.org/abstract/document/8822482.

Zhang, F., Chen, J., Mao, T., & Wang, Z. (2021). Feedback interval optimization for MISO LiFi systems. IEEE Access, 9, 136811-136818.https://ieeexplore.ieee.org/abstract/document/9557310.