RESEARCH IDEA(S)

Faculty of Sciences

Department of Computer Science and Information Systems

COMBINING SELECTED HYPERPARAMETER OPTIMISERS FUNCTIONALITY (CREATING A NEW ALGORITHM) TO IMPROVE DEEP LEARNING MODELS

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

1. INTRODUCTION

Hyperparameter optimization (HPO) 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). HPO involves the search for optimal hyperparameters governing a model's architecture, regularization, and optimization strategy, while fine-tuning entails adapting pre-trained models to specific tasks or domains, leveraging transfer learning to enhance performance on target datasets (Zhu et al., 2019).

The goal of Machine Learning applications is to produce the most accurate prediction as possible from a given dataset. In doing this, the Machine Learning model needs to be preprocessed in a way that will result in true positive outcomes. Hyper-parameter Tuning optimisation and fine-tuning are processed in the model to achieve greater performance and have a great impact on the performance of the model in any sort of application.

2. SECTION 2. SCIENTIFIC BACKGROUND OF THE TARGETED SUBJECT

Hyper-parameter refers to parameters that cannot be updated during the training of machine learning. They can be involved in building the structure of the model, such as the number of hidden layers and the activation function, or in determining the efficiency and accuracy of model training, such as the learning rate (LR) of stochastic gradient descent (SGD), batch size, and optimizer (hyp).The history of HPO dates back to the early 1990s (Ripley, 1993; 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): 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).

3. FUNDAMENTAL / APPLIED PROBLEM ADDRESSED BY THE PROPOSED RESEARCH

Despite extensive research in both areas, there is a significant gap regarding their integration and combined effects. To the best of my knowledge, from previous research, hyperparameter optimization methods and fine-tuning methods have been compared to one another as in a study by Liu and Wang (2021). In other cases, hyper-parameter optimisation techniques have been the sole focus to find a set of hyperparameter values which gives us the best model for our data in a reasonable amount of time as is the case with Bischl et al., (2023).

Comprehensive studies systematically investigating the integration of HPO and fine-tuning methods could unveil synergistic effects and optimal combinations, enhancing model performance across various domains (Snoek et al., 2012). Research efforts aimed at developing scalable and efficient optimization
algorithms, capable of handling large-scale datasets and complex model architectures, are crucial for real-world applications of machine learning. Investigating the impact of hyperparameter choices on fine-tuning outcomes, particularly in domains with limited labeled data or significant domain shifts, could provide valuable insights into improving model generalization and transferability (Bender et al., 2018).

4. MAIN OBJECTIVES

1. Identify hyperparameter optimization methods to be used.
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. Identify and experiment the combined optimisers on, particularly Convolutional Neural Network (CNN), Radial basis function (RBF) and Long Short-Term Memory (LSTM).

5. SPECIFIC OBJECTIVES

1. Create a novel algorithm which can be used as part of the existing optimisers.

6. METHODOLOGY PROPOSED TO ACCOMPLISH THE OBJECTIVES

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.

7. SELECTED REFERENCES

Liu, X., & Wang, C. (2021). An empirical study on hyperparameter optimization for fine-tuning pre-trained language models. arXiv preprint arXiv:2106.09204.

Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., ... & Lindauer, M. (2023). Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 13(2), e1484.

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).

Zhu, X., Lei, J., Liu, H., Shi, Z., & Li, S. Z. (2019). Deep learning for generic object detection: A survey. arXiv preprint arXiv:1901.07755.

Bender, G., Kindermans, P. J., Zoph, B., Vasudevan, V., & Le, Q. V. (2018). Understanding and simplifying one-shot architecture search. In International Conference on Machine Learning (pp. 550-559).

8. OTHER RESEARCH IDEAS INCLUDE

1. Changing Machine Learning algorithms with an HPO
2. Creating a evaluating HPO framework
3. Improved existing HPO - creating a new algorithm