BRNO UNIVERSITY OF TECHNOLOGY FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2023

http://excel.fit.vutbr.cz

Design of Accuracy Predictors for CNNs

Šimon Šmída*

Abstract

This paper focuses on improving the efficiency of Neural Architecture Search (NAS) by predicting the performance of Convolutional Neural Networks (CNN) before training. A performance predictor is proposed, using machine learning methods like linear regression, random forest, and Multi-layer Perceptron (MLP) to estimate the classification accuracy of new CNN architectures. The study emphasizes the importance of feature selection, extraction, and representation of neural architectures, as well as the analysis and comparison of different predictor approaches. The proposed methodology efficiently predicts the performance of new CNN architectures, speeding up the NAS process and achieving competitive accuracy with existing methods. This research contributes to the understanding of the underlying factors influencing architecture performance and is valuable for guiding the NAS process towards optimal architectures more efficiently, saving computational resources and time, thus benefiting both science and practical applications.

*xsmida03@vutbr.cz, *Faculty of Information Technology, Brno University of Technology*

1. Introduction

[Motivation] The efficiency of Neural Architecture Search (NAS) is critical in finding optimal architectures for Convolutional Neural Networks (CNN). However, the search process often requires significant computational resources and time. To address this issue, the analysis of performance predictors that can estimate the classification accuracy of new CNN architectures without training them is proposed.

[Problem definition] The goal is to develop, analyze and compare performance predictors that can accurately estimate the classification performance of CNN architectures, guiding the NAS process more efficiently and ultimately reducing the computational resources and time required.

[Existing solutions] Current NAS methods explore the search space typically using reinforcement learning, evolutionary algorithms, or gradient-based approaches. However, these methods often demand high computational costs [1]. To alleviate this issue, various performance estimation strategies have been developed, which can be grouped into different families, such as model-free and model-based estimators [2]. Model-free performance predictors employ methods like learning curve extrapolation, one-shot techniques, or other heuristics, while model-based predictors employ machine-learning techniques to estimate model performance. In this work, the focus is on model-based predictors, as they have shown promising results in reducing computational costs and search time while maintaining high performance (low query time) [3].

[Our solution] Our approach involves using machine learning methods, such as linear regression, random forest, and Multi-layer Perceptron (MLP), to develop performance predictors that estimate the classification accuracy of new CNN architectures. The predictors are trained on a dataset of previously evaluated architectures and their corresponding classification accuracies (e.g., `NAS-Bench-101` [4]). We emphasize the importance of feature selection, extraction, and neural architectures representation in building effective predictors. Our research also includes a comprehensive analysis of different feature representations and their impact on predictor performance. Additionally, we compare various predictor approaches to understand their respective strengths and weaknesses, contributing to a better understanding of the factors influencing the efficiency of the NAS process.

[Contributions] The research contributes to the understanding and analysis of performance predictors in the context of estimating the performance of new CNN architectures, leading to a more efficient NAS process. The emphasis is on the importance of feature selection and investigating the impact of different feature representations on predictor accuracy. By comparing methods flattening the network architecture structure as features and methods preserving structural information, we provide insights into their respective advantages and limitations. Our work offers a comprehensive analysis of existing techniques, exploring their effectiveness and potential improvements, ultimately advancing the field of performance prediction and NAS efficiency.

2. Performance Prediction for NAS

In this work, we develop a performance predictor that employs machine learning methods to estimate the classification accuracy of new CNN architectures. We train the predictor on a dataset of previously evaluated architectures and their corresponding classification accuracies, enabling it to generalize and make predictions for unseen architectures. *Figure 1* illustrates the typical process of building a performance predictor, which includes feature extraction, architecture representation, and predictor training and evaluation.

To optimize the predictor models, a hyperparameter tuning strategy that systematically searches for the best combination of hyperparameters for each machine learning method is employed. This strategy improves the performance of the predictor models, ensuring that they are optimally configured for the specific task at hand.

3. Applying the Predictor

Utilizing the trained predictor models, as illustrated in *Figure 2* inspired by Google Brain's Neural Predictor [5], predictions for new CNN architectures can be made efficiently. This approach demonstrates the application of the performance predictor, starting from the search space with many random architectures, through the regression model (predictor), to the top K selection, training, and validation, ultimately picking the best validation accuracy. The approach allows for the effective estimation of the performance of previously unexplored architectures, ultimately guiding the NAS process more effectively in the search for optimal structures.

4. Conclusions

The focus of this work was the process of developing performance predictors to effectively estimate the classification accuracy of new CNN architectures, enabling a more efficient NAS process. By employing machine learning methods and hyperparameter tuning, the predictors achieve similar accuracy to existing NAS methods while reducing computational cost and time. The importance of feature selection and extraction, as well as the representation of neural architectures in building effective predictors was emphasized. The research provides a comprehensive analysis of different feature representations and their impact on predictor performance. The comparison of various predictor approaches was conducted to understand their respective strengths and weaknesses, contributing to a deeper understanding of the factors that influence the efficiency of the NAS process. *Figure 3* shows a plot of the predicted versus actual accuracy values (in percentages) for the Random Forest predictor, highlighting its performance in estimating the classification accuracy of new CNN architectures. This work paves the way for further advancements in NAS by shedding light on the critical role of feature selection, extraction, and architectural representation.

Acknowledgements

I would like to thank my supervisor prof. Ing. Lukáš Sekanina. Ph.D. for his mentoring and help during the creation of this work.

References

[1] Frank Hutter, Lars Kotthoff, and Joaquin Vanschoren, editors. *Automatic Machine Learning: Methods, Systems, Challenges*. Springer, 2019.

[2] Thomas Elsken, Jan Hendrik Metzen, and Frank Hutter. Neural architecture search. In Hutter et al. [1], pages 69–86.

[3] Colin White, Arber Zela, Binxin Ru, Yang Liu, and Frank Hutter. How powerful are performance predictors in neural architecture search? *ArXiv*, abs/2104.01177, 2021.

[4] Chris Ying, Aaron Klein, Eric Christiansen, Esteban Real, Kevin Murphy, and Frank Hutter. NAS-bench-101: Towards reproducible neural architecture search. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 7105–7114, Long Beach, California, USA, 09–15 Jun 2019. PMLR.

[5] Wei Wen, Hanxiao Liu, Hai Helen Li, Yiran Chen, Gabriel Bender, and Pieter-Jan Kindermans. Neural predictor for neural architecture search. *ArXiv*, abs/1912.00848, 2019.