GA-based Training-Free NAS Algorithm with Hybrid Score Function

Hsieh Cheng-Han

I. Abstract

Most neural architecture searches (NASs) are time-consuming caused by the fact that, during the searching, a candidate architecture must be trained to evaluate how good of this architecture. This is why some of training-free NAS algorithms have been proposed in recent years.

Although the training-free NASs are typically faster than training-based NAS method, however, the correlation between score value and the result of an architecture is not well enough in most cases.

To address this problem, we propose a genetic-based training-free NAS algorithm with hybrid training-free score function, which combines three highly heterogeneous training-free score functions to evaluate an architecture. In this method, the genetic algorithm plays a role to guide the searches of NAS algorithm while the hybrid training-free score function plays the role to evaluate a new candidate architecture during the convergencr process of GA. More precisely, the first score function is noise immunity for neural architecture search without search (NINASWOT), as an evaluation of pattern recognition ability, second one is maximum-entropy detection (MAE-DET), as an evaluation of the entropy of an architecture and the third one is condition number of neural tangent kernel (NTK), as an evaluation of the speed of converge.

To evaluate the performance of the proposed algorithm, we compared it with several NAS algorithms, including weight-sharing methods, non-weight-sharing methods, and neural architecture search without training (NASWOT). We expect develope a faster and more accurate training-free NAS algorithm.

II. Introduction

Neural architecture search (NAS) has recently drawn a big amount of attention, since the ability to automatically design a ”good” neural architecture. By leveraging machine learning algorithms (Zoph & Le, [1]), NAS algorithms can explore a search space, which is comprised of numerous potential architectures, to find out a good architectures that outperform those designed by human experts. In recent years, the use of NAS is widespread, from object detection (Sun et al., [2]), image recognition (Mellor et al., [3]) and speech recognition (Zheng et al., [4], Mehrotra et al., [5]) to natural language processing(Jiang et al., [6], Klyuchnikov et al., [7], Wang et al., [8]).

Despite the promising results of NAS, there are still many challenges to conquer. One of the major problems is the extremely high computational cost to search for an optimal architecture, which can make NAS impractical for real-world applications, particularly on resource-constrained platforms like embedded system. The reason why NAS is costly is that during the searching, a candidate architecture must be trained to evaluate how good of this architecture.

To overcome this challenge, recent works developed and proposed lots of method which is so called training-free NAS. Some use mean absolute error random sampling (MRS) (Camero et al., [9], Camero et al., [10], Camero et al., [11]) as a training-free score function and some use neural tangent kernel (NTK) (Chen et al., [12], Wang et al., [13], Shu et al., [14]), while some use gradient-based methods (Zhang & Jia, [15], Cavagnero et al., [16], Abdelfattah et al., [17]). However, most of score functions suffer from low correlation between score value and the result of an architecture, leading to a trap that no matter how good the search method is used, we can hardly find an optimal architecture. The major problem causes the low correlation is that a single score function can only evaluate one perspect/characteristic of an architecture. By cooperating three heterogeneous score functions with genetic-based search method, we shall evaulate an architecture from different aspects. More specifically, the first score function is noise immunity for neural architecture search without search (NINASWOT) (Wu et al. [18]). Based on the work of Mellor et al., they proposed the measurement of the correlation between binary activation paterns of input data at each ReLU layer. Later, Wu et al. found a high score obtained by such a function may not correspond to a high-performence model. Thus, they additionally applied noisy immunity method, called NINASWOT. Second one is maximum-entropy detection (MAE-DET) (Sun et al. [2]). Based on the work of Sun et al., they proposed to find the maximum of the differential entropy by calculating the variance of the output at the final layer with the input Gaussian noise images. Third one is condition number of neural tangent kernel (NTK) (Chen et al. [12]), which is defined as

$$\mathcal{K}_{\mathcal{N}} = \frac{\lambda_{\max}(\hat{\Theta})}{\lambda_{\min}(\hat{\Theta})} \tag{1}$$

Chen et al. used the condition number of NTK as the score function to estimate the performance of a network. It’s shall find good architectures in reasonable time and computational resource. The major contribution of this paper can be summarized as follow:

-

The remainder of this paper is organized as follows:

REFERENCES

[1] B. Zoph and Q. V. Le, "Neural architecture search with reinforcement learning," 2016. [Online]. Available: https://arxiv.org/abs/1611.01578

[2] Z. Sun, M. Lin, X. Sun, Z. Tan, H. Li, and R. Jin, "Mae-det: Revisiting maximum entropy principle in zero-shot nas for efficient object detection," 2021. [Online]. Available: https://arxiv.org/abs/2111.13336

[3] J. Mellor, J. Turner, A. Storkey, and E. J. Crowley, "Neural architecture search without training," 2020. [Online]. Available: https://arxiv.org/abs/2006.04647

[4] H. Zheng, K. An, and Z. Ou, "Efficient neural architecture search for end-to-end speech recognition via straight-through gradients," 2020. [Online]. Available: https://arxiv.org/abs/2011.05649

[5] A. Mehrotra, A. G. C. P. Ramos, S. Bhattacharya, Ł. Dudziak, R. Vipperla, T. Chau, M. S. Abdelfattah, S. Ishtiaq, and N. D. Lane, "{NAS}-bench-{asr}: Reproducible neural architecture search for speech recognition," in *International Conference on Learning Representations*, 2021. [Online]. Available: https://openreview.net/forum?id=CU0APx9LMaL

[6] Y. Jiang, C. Hu, T. Xiao, C. Zhang, and J. Zhu, "Improved differentiable architecture search for language modeling and named entity recognition," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 3585–3590. [Online]. Available: https://aclanthology.org/D19-1367

[7] N. Klyuchnikov, I. Trofimov, E. Artemova, M. Salnikov, M. Fedorov, and E. Burnaev, "Nas-bench-nlp: Neural architecture search benchmark for natural language processing," 2020. [Online]. Available: https://arxiv.org/abs/2006.07116

[8] H. Wang, Z. Wu, Z. Liu, H. Cai, L. Zhu, C. Gan, and S. Han, "Hat: Hardware-aware transformers for efficient natural language processing," 2020. [Online]. Available: https://arxiv.org/abs/2005.14187

[9] A. Camero, J. Toutouh, and E. Alba, "Low-cost recurrent neural network expected performance evaluation," 2018. [Online]. Available: https://arxiv.org/abs/1805.07159

[10] A. Camero, H. Wang, E. Alba, and T. Bäck, "Bayesian neural architecture search using a training-free performance metric," *Applied Soft Computing*, vol. 106, p. 107356, jul 2021. [Online]. Available: https://doi.org/10.1016%2Fj.asoc.2021.107356

[11] A. Camero, J. Toutouh, and E. Alba, "Reliable and fast recurrent neural network architecture optimization," 2021. [Online]. Available: https://arxiv.org/abs/2106.15295

[12] W. Chen, X. Gong, and Z. Wang, "Neural architecture search on imagenet in four gpu hours: A theoretically inspired perspective," 2021. [Online]. Available: https://arxiv.org/abs/2102.11535

[13] H. Wang, Y. Wang, R. Sun, and B. Li, "Global convergence of maml and theory-inspired neural architecture search for few-shot learning," 2022. [Online]. Available: https://arxiv.org/abs/2203.09137

[14] Y. Shu, S. Cai, Z. Dai, B. C. Ooi, and B. K. H. Low, "Nasi: Label- and data-agnostic neural architecture search at initialization," 2021. [Online]. Available: https://arxiv.org/abs/2109.00817

[15] Z. Zhang and Z. Jia, "Gradsign: Model performance inference with theoretical insights," 2021. [Online]. Available: https://arxiv.org/abs/2110.08616

[16] N. Cavagnero, L. Robbiano, B. Caputo, and G. Averta, "Freerea: Training-free evolution-based architecture search," 2022. [Online]. Available: https://arxiv.org/abs/2207.05135

[17] M. S. Abdelfattah, A. Mehrotra, L. Dudziak, and N. D. Lane, "Zero-cost proxies for lightweight nas," 2021. [Online]. Available: https://arxiv.org/abs/2101.08134

[18] M.-T. Wu, H.-I. Lin, and C.-W. Tsai, "A training-free genetic neural architecture search," in *Proceedings of the 2021 ACM International Conference on Intelligent Computing and Its Emerging Applications*, ser. ACM ICEA '21. New York, NY, USA: Association for Computing Machinery, 2022, p. 65–70. [Online]. Available: https://doi.org/10.1145/3491396.3506510