Improving scores of the NetHack Challenge by leveraging transformer architecture

Tergel Munkhbat[1,2], Busueng Cho[1,2]

[1]Korea Institute of Science and Technology Information, Korea

[2]University of Science & Technology, Korea

tergelmunkhbat@kisti.re.kr, bscho@kisti.re.kr

Abstract

The NetHack Challenge is a competition where teams compete to construct optimal agents using reinforcement learning (RL) or other algorithms in the NetHack Learning Environment (NLE). Numerous researchers have built outstanding agents to play the game and achieved better than the baseline scores. The objective of this study is to reach higher scores than the baseline scores by proposing two types of models. The first model incorporates the Dueling network architecture, and the second model connects the local and global information by modifying the transformer architecture. The experimental results demonstrate that the proposed transformer architecture performs better than the baseline even though it is trained for less time. Moreover, the early results strongly confirm that the architecture is capable of addressing the NetHack challenge. This indicates that the modified transformer architecture can be regarded as an efficient model for this challenge.

Ⅰ. Introduction

The NetHack Challenge [1] is a competition, and it plays a crucial role in the development of efficient reinforcement learning (RL) algorithms since the NetHack Learning Environment (NLE) is extremely challenging for both humans and state-of-the-art RL models. NetHack includes hundreds of distinct creatures and intricate environment dynamics. The objective of an agent is to retrieve the amulet meanwhile it should collect magical items and enough food to survive.

Many researchers have developed outstanding agents to play the game and achieved better than the baseline scores [2]. Their results suggested that symbolic agents currently outscore much more than deep RL, and no agent has come close to winning the game so far [3], meaning that the NetHack Challenge is a long-term benchmark for RL research.

This study focuses on two deep RL algorithms to achieve higher scores than the baseline in the NLE. Dueling networks and modified versions of transformer architectures are designed to improve the scores. Furthermore, preliminary findings indicate that the architectures are capable of solving the NetHack problem. Hence, the modified transformer design can be considered an efficient model for this task.

Ⅱ. Methodology

The current investigation involved the Dueling network architecture [4] and the transformer architecture [5], which were used to modify the baseline code provided in the challenge.

1. Dueling network architecture

This architecture separates estimates of the value and advantage functions while they share standard convolutional filters. Moreover, particular aggregating layers combine these two streams to generate an optimal state-action value Q function shown in Figure. 1

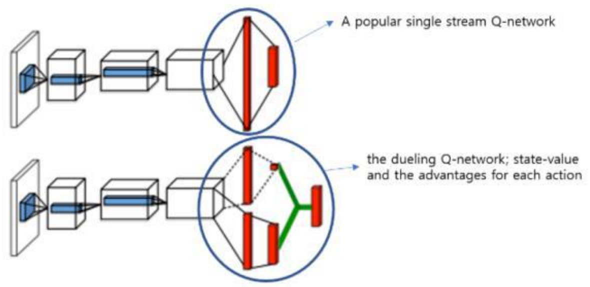

Figure 1 Structure of Dueling network [4]: Q-network (top) and the dueling Q-network (bottom)

2. Transformer for Encoder

It can be observed that the correlation between local and global information is crucial to solving the NetHack challenge. Therefore, this work connects local and global information by modifying the transformer architecture. To this end, it considers the global visual features from CNN as query and local visual context as key and value. Moreover, to reduce the computational cost in the transformer, the depth-wise separate convolution used in MobileNet [6] was utilized. The details can be seen in Figure. 2. This model extracts the final feature by passing both the message and the visual context through the MLP, and it adopts a sequential structure through the LSTM at every stage.

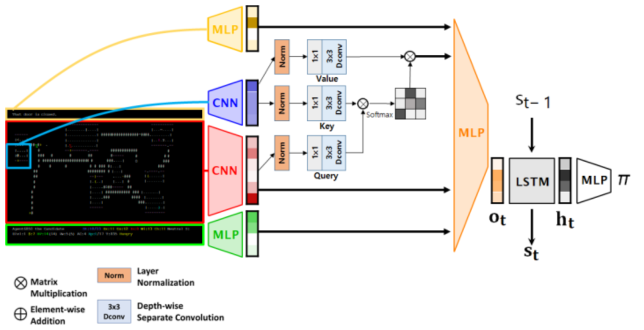

Figure 2 Structure of the transformer architecture

Ⅲ. Experiments

The experiments were conducted on these two proposed models. Both models were trained with batch size 16. Due to the lack of GPU resources, both models have not been trained until they converge. Generally, networks in the NetHack challenge need to train for 0.5 billion or more iterations, but it takes more than two weeks with a single GPU. Therefore, this study reports the results of 0.2 billion iterations for the baseline and the dueling architecture; and the results of 0.05 billion iterations for the transformer architecture. Moreover, it takes a long time to evaluate. Hence, the evaluation was conducted for 32 episodes instead of 512 episodes.

Ⅳ. Results

Table 1 displays the results of each experimental setting. It can be observed that the time to complete this task and good performance are not significantly related. For example, it takes less time even though the models cannot solve the challenge well because it moves on to the next episode quickly. Interestingly, the modified transformer architecture reaches higher than the baseline scores even though it is trained for fewer iterations. Hence, if the proposed architecture is trained enough, it can perform much better than the baseline.

Table 1 Results of experiments

	Baseline	Dueling	Transformer
Median	78.0	91.5	**145.5**
Mean	**126.92**	87.3	125.17
Test Time	12.29m	**11.17m**	39.03m
Train Iteration	0.2 billion	0.2 billion	0.05 billion

Ⅴ. Conclusion

The modified transformer architecture has demonstrated the effectiveness of solving the NetHack challenge because it can efficiently correlate local and global information. More specifically, the proposed model accomplishes better than the baseline scores even though it is trained for fewer iterations. In this report, the early results indicate that the proposed architectures are able to address the NetHack challenge. However, due to the limitation of hardware resources, CNN architecture was used for representation. Graph representation might improve the results instead of CNN layers. Moreover, more advanced DQN-based RL architectures might achieve higher scores than the proposed architecture. Hence, the future work should include follow-up work to enhance and develop optimal agents of the NetHack challenge.

ACKNOWLEDGMENT

We would like to express our gratitude to the KREONet project for the Development and Establishment of National Large-scale Data Exchange Node. We also extend our thanks to all individuals who have directly or indirectly supported this work.

REFERENCES

[1] https://nethackchallenge.com/

[2] https://www.aicrowd.com/challenges/neurips-2021-the-nethack-challenge/leaderboards

[3] Hambro, Eric, et al. "Insights from the neurips 2021 nethack challenge." NeurIPS 2021 Competitions and Demonstrations Track. PMLR, 2022.

[4] Wang, Ziyu, et al. "Dueling network architectures for deep reinforcement learning." International conference on machine learning. PMLR, 2016.

[5] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).

[6] Howard, Andrew G., et al. "Mobilenets: Efficient convolutional neural networks for mobile vision applications." arXiv preprint arXiv:1704.04861 (2017).