

Generalized Representation Learning Methods for Deep Reinforcement Learning

Hanhua Zhu

GSII, The University of Tokyo, Japan
 zhu-hanhua@g.ecc.u-tokyo.ac.jp

Abstract

Deep reinforcement learning (DRL) increases the successful applications of reinforcement learning (RL) techniques but also brings challenges such as low sample efficiency. In this work, I propose generalized representation learning methods to obtain compact state space suitable for RL from a raw observation state. I expect my new methods will increase sample efficiency of RL by understandable representations of state and therefore improve the performance of RL.

1 Introduction

Deep reinforcement learning (DRL), which combines reinforcement learning (RL) with deep neural networks, sheds light on the viability of training directly from raw observations such as images. This advancement increases the successful applications of RL techniques but also brings challenges. One is low sample efficiency since considerable interactions are needed for learning compact representations of the raw observation space. This property limits the performance of DRL, particularly in real-world tasks such as robotic control where collecting interaction data is expensive and time-consuming.

An extensively studied and effective method to increase sample efficiency is to conduct representation learning to obtain compact state space for RL. For model-free RL, some prior works are end-to-end methods and we can understand that they implicitly learn representation as a part of policy learning [Mnih *et al.*, 2015] whereas a number of prior other works focus on using a distinct representation learning procedure [Nair *et al.*, 2018; Lee *et al.*, 2019]. On the other hand, for model-based RL, representation learning plays an important role in constructing a predictive model [Zhang *et al.*, 2018].

In this work, I propose generalized representation learning methods to obtain compact state space suitable for RL from a raw observation state. I expect my new methods will increase sample efficiency of RL by understandable representations of state and therefore improve the performance of RL, especially in tasks where interacting with environments are expensive.

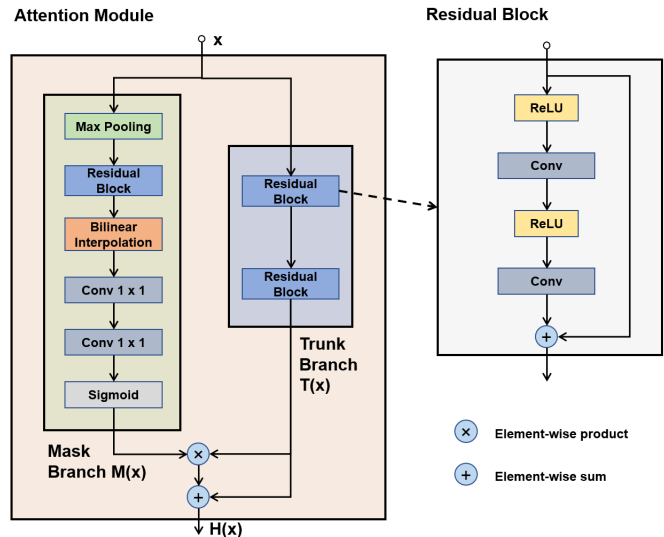


Figure 1: The structure of the attention module.

2 Deep Residual Attention Reinforcement Learning

I have investigated the feasibility of incorporating attention mechanisms into DRL to increase its ability to learn a good representation from visual observations and proposed an end-to-end model called Deep Residual Attention Reinforcement Learning (DRARL) [Zhu and Kaneko, 2019]. DRARL solves the representation learning task and the policy learning task together and outperforms its counterpart Importance Weighted Actor-Learner Architecture (IMPALA) [Espeholt *et al.*, 2018] in several Atari games. In this section, I am going to describe the proposed method DRARL.

2.1 Method

DRARL uses the attention module shown in Fig.1 as a basic unit to construct the model. An attention module processes input with two branches similar to [Wang *et al.*, 2017]. The trunk branch is responsible for the policy learning, doing the same as the network in other end-to-end DRL. On the other hand, the mask branch computes soft weights on input features, which aims to help the trunk branch learn a better representation of input by increasing the weights of crucial fea-

Game	IMPALA	DRARL
Amidar	757.45	749.50
Breakout	473.00	497.29
Centipede	6716.95	6899.73
Krull	5966.99	8053.78
Seaquest	1631.57	2280.00

Table 1: The score of agents trained by DRARL and original IMPALA

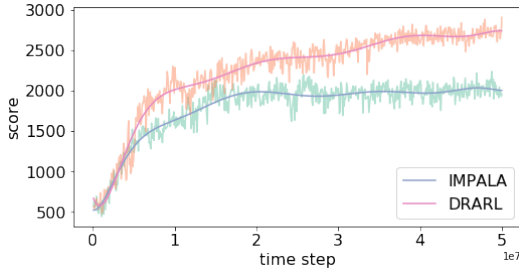


Figure 2: Comparison of DRARL and IMPALA in Krull. (The score is computed in training environment where agents only have single life.)

tures while decreasing the weights of irrelevant features. The output of an attention module is computed by:

$$H_{i,c}(x) = (1 + M_{i,c}(x)) * T_{i,c}(x), \quad (1)$$

where i ranges over all input features and c is the index of the channel in convolutional neural networks. I choose this structure mainly because of its simplicity. The additional structure is implemented while the hyperparameters and the loss functions are kept the same. Furthermore, the model enjoys the benefits of representation learning at the cost of only slightly more parameters. With the structure shown in Fig. 1, DRARL has only four percent more parameters than IMPALA. The whole model is trained in the same way as IMPALA which is a distributed computation efficient DRL algorithm.

2.2 Result

I have evaluated DRARL in different environments and the results of Atari games are shown in Table 1. More results and details of the model are published in [Zhu and Kaneko, 2019]. From Table 1, we can find that benefiting from obtaining good representations, DRARL outperforms the original IMPALA in four games, especially in Krull and Seaquest. Although no improvement is observed in Amidar, no significant negative effect is brought about by the additional attention mechanism either.

3 Discussion

In this work, I plan to propose representation learning methods to obtain compact state space from raw observations to improve sample efficiency of DRL. For this purpose, I have proposed DRARL, which has demonstrated its superior sample efficiency in some Atari games, but it does have limitations. First, DRARL learns representations deterministically, which means it has difficulty representing the uncertainty in

an environment. Second, DRARL is only applied in model-free RL and no effort has been put into model-based RL.

In the next step, I will propose methods to solve these limitations. For the first limitation, previous works have observed that constructing a stochastic latent variable model is effective to learn representations of the uncertainty in an environment [Doerr *et al.*, 2018; Lee *et al.*, 2019]. This induces us to research the stochastic latent variable model in the future. Although the distinct stochastic latent variable model will be more complex than the representation learning model used in DRARL, better representations obtained by the stochastic latent variable model will considerably improve sample efficiency and the final performance, which has been demonstrated in prior research [Lee *et al.*, 2019].

As the goal of this research is to propose generalized representation learning methods for RL, model-based RL and model-free RL both should be considered. The latent variable model has been used to obtain representation for both model-free RL [Lee *et al.*, 2019] and model-based RL [Zhang *et al.*, 2018], which attracts us to do further experiments on this method.

References

- [Doerr *et al.*, 2018] Andreas Doerr, Christian Daniel, and Martin Schiegg *et al.* Probabilistic recurrent state-space models. *arXiv preprint arXiv:1801.10395*, 2018.
- [Espelholt *et al.*, 2018] Lasse Espelholt, Hubert Soyer, and Remi Munos *et al.* Impala: Scalable distributed deep-rl with importance weighted actor-learner architectures. *arXiv preprint arXiv:1802.01561*, 2018.
- [Lee *et al.*, 2019] Alex X Lee, Anusha Nagabandi, Pieter Abbeel, and Sergey Levine. Stochastic latent actor-critic: Deep reinforcement learning with a latent variable model. *arXiv preprint arXiv:1907.00953*, 2019.
- [Mnih *et al.*, 2015] Volodymyr Mnih, Koray Kavukcuoglu, and David Silver *et al.* Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.
- [Nair *et al.*, 2018] Ashvin V Nair, Vitchyr Pong, and Mur-taza Dalal *et al.* Visual reinforcement learning with imagined goals. In *Advances in Neural Information Processing Systems*, pages 9191–9200, 2018.
- [Wang *et al.*, 2017] Fei Wang, Mengqing Jiang, and Chen Qian *et al.* Residual attention network for image classification. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6450–6458. IEEE, 2017.
- [Zhang *et al.*, 2018] Marvin Zhang, Sharad Vikram, and Laura Smith *et al.* Solar: deep structured representations for model-based reinforcement learning. *arXiv preprint arXiv:1808.09105*, 2018.
- [Zhu and Kaneko, 2019] Hanhua Zhu and Tomoyuki Kaneko. Deep residual attention reinforcement learning. In *2019 International Conference on Technologies and Applications of Artificial Intelligence (TAAI)*, pages 1–6. IEEE, 2019.