Domain Adaptation for Reinforcement Learning on the Atari

Extended Abstract

Thomas Carr
Computer Science, Aston University
Birmingham, United Kingdom
carrtp@aston.ac.uk

Maria Chli
Computer Science, Aston University
Birmingham, United Kingdom
m.chli@aston.ac.uk

George Vogiatzis
Computer Science, Aston University
Birmingham, United Kingdom
g.vogiatzis@aston.ac.uk

ABSTRACT

Deep Reinforcement learning is a powerful machine learning paradigm that has had significant success across a wide range of control problems. This success often requires long training times to achieve. Observing that many problems share similarities, it is likely that much of the training done could be redundant if knowledge could be efficiently and appropriately shared across tasks. In this paper we demonstrate a novel adversarial domain adaptation approach to transfer state knowledge between domains and tasks on the Atari game suite. We show how this approach can successfully transfer across very different visual domains of the Atari platform. We focus on semantically related games that involve returning a ball with the user controlled agent. Our experiments demonstrate that our method reduces the number of samples required to successfully train an agent to play an Atari game.

KEYWORDS

Deep Learning; Reinforcement Learning; Domain Adaptation

ACM Reference Format:
Thomas Carr, Maria Chli, and George Vogiatzis. 2019. Domain Adaptation for Reinforcement Learning on the Atari. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019,* IFAAMAS, 3 pages.

1 INTRODUCTION

Deep Reinforcement Learning (DRL) successfully extends the reinforcement learning paradigm to complex control problems by alleviating the need for expert hand-crafted features and allowing for end-to-end learning from the input space, for example from images. One way to view learning within DRL is to consider the hidden layers as learning a state representation on top of which a policy can be learned. This view has held true in the standard deep learning paradigm, where the learned features are often repurposed through direct transfer with or without fine-tuning for a subsequent task [11]. Learning the feature space from a sparse reward signal is difficult and so deep RL algorithms typically require a large number of training samples; this is further exacerbated by the standard RL problem of the exploration-exploitation trade-off.

Transfer Learning can help by taking advantage of previous training to reduce the effort needed to solve a new task. In this work we show how learning an input mapping from source to target task embedding in an unsupervised manner can be used as an initialization step to improve learning even when the input domain, task and action space vary.

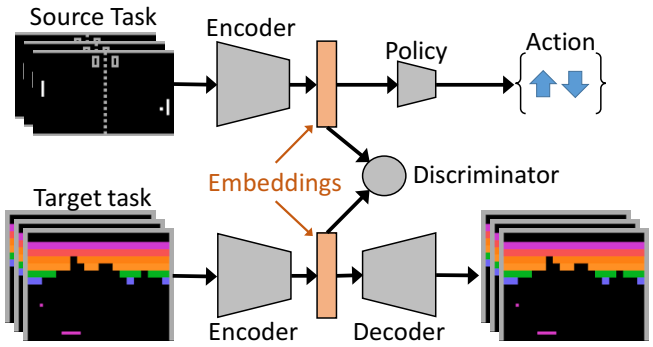

Figure 1: Our agent's architecture for domain adaptation, combining adversarially discriminative domain adaptation and adversarial autoencoder architectures. A fully-trained source task is depicted on top, while on the bottom is the target task autoencoder.

2 ARCHITECTURE AND METHOD

We propose an Adversarial Domain Adaptation approach to the problem of Knowledge Transfer for reinforcement learning in the Atari domain. In this view the source and target problem share a state space and we wish to learn a mapping from the target observations to this previously-learned source state space. Complicating this problem is the lack of an obvious alignment between source and target observation pairs.

Our proposed Architecture, depicted in Fig. 1, builds on the Adversarial AutoEncoder (AAE) [6]. The AAE combines an autoencoder (AE) and generative adversarial network (GAN) [4]. The GAN regularizes the embedding space of the AE as an alternative approach to the approach of the variational autoencoder [5] In our approach we replace the standard Gaussian distribution with samples from a learned policy embedding.

By using samples from the source task and the pre-trained source policy network, we can generate samples from the embedding of the source policy, which we use to represent the state space of our policy. In this way, real states are sampled from the source domain. We train a generator function from target observations to match the distribution of source states, as learned by our source policy. If the tasks are related and require similar state features pre-training the generator in this manner should make it easier to learn a new task, as the problem of learning a state descriptor

Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 13–17, 2019, Montreal, Canada.

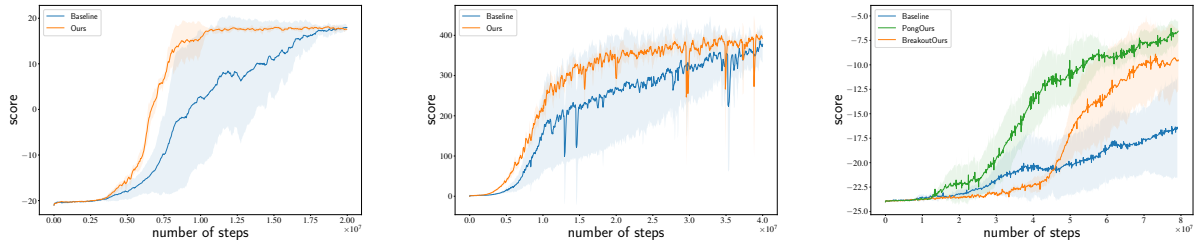

Figure 2: The 100-episode average score plot against total number of steps. From left to right, 1: Source Task: Breakout, Target Task Pong. 2: Source Task: Pong, Target Task: Breakout. 3: Source Tasks: Pong, Breakout, Target Task: Tennis.

has been addressed. To ensure the learned mapping maintains all the information contained in the target observations an AE is also trained in conjunction with the GAN element of the architecture.

An alternative view is to see the source embedding as a regularizing force on the auto-encoder embedding. This interpretation aligns the method with alternative approaches many of which seek to minimize the Kullback-Liebler divergence between aligned source and target embeddings [12]. Shared embeddings are a natural way to conceptualize transfer as we consider related domains to be part of a shared representation space. In this case the GAN element allows us to perform that minimization without direct correspondences. Applying a GAN to align reinforcement learning domains for generalization has also recently been applied in the robotics context in [10].

3 EXPERIMENTS AND RESULTS

Our experiments investigate the applicability of Adversarial Domain Adaptation to the reinforcement learning problem. We use the Arcade Learning Environment (ALE) [2] to provide a set of arcade games which are commonly used as benchmarks for DRL. We interface with ALE through the OpenAi Gym platform [3]. Our experiments focus on transfer between pairs of games. We begin by selecting pairs of games which we deem to be related in order to verify the effectiveness of our approach; we leave automatically identifying appropriate source tasks for future work. In this work we focus on three games: Pong, Breakout and Tennis.

To perform transfer we train an agent to solve the source task, in our case using A2C [1, 9]; a type of actor-critic algorithm. We then run this trained policy and store samples from the source task; in our case 100000 images. We also run a random policy on the target task and store 100000 sample images from this domain as well. We now sample from these datasets to train our Adversarial AutoEncoder framework. The generator weights are then used to initialise the weights of a new policy network for the target task. The final layers for the policy and value function are initialised randomly, as they were untrained during the pre-training phase and the action space differs from the source task's action-space.

We can see in Fig. 2 that our approach significantly improves learning on the target tasks. The baseline we compare against is training a standard randomly initialised A2C agent on only the target task. The comparison clearly shows that our approach has much promise.

Since we always start training with randomly initialised policy and value function layers, we cannot expect transfer to provide a jumpstart [8] in rewards. We do however see observe that initialising the state representation can improve over tabula rasa learning by decreasing the number of samples required to learn a solution and therefore our method does achieve faster learning, which suggests our learned representation provides a better initialisation than learning tabula rasa.

When examining the result of transfer to Tennis from either Pong or Breakout we can see that while both games provide benefit there is a significant difference between them. Conceptually Pong and Tennis are more aligned as they are both versions of the same game with a similar scoring mechanic. Breakout however is rotationally aligned with Tennis, both games play vertically on the screen, pong plays horizontally. Understanding this difference and developing a method to identify the best sources for transfer is left as future work.

4 CONCLUSION

This paper presents an adversarial method for knowledge transfer in reinforcement learning. We have demonstrated how this approach can be used for domain adaptation to improve performance on the difficult task of learning to play Atari games. As evident from our results, the technique can help learning in the reinforcement learning case, even when the final layer needs to be relearned as the action space and task have changed, making this application of domain adaptation significantly more difficult than standard applications of the technique.

Transfer Learning is a complex problem with many approaches and applications from reducing the number of samples required to solve a problem to improving the usefulness of simulation for training of agents. Future work will compare our approach to other Transfer learning methods such as the Progressive Neural Network [7].

REFERENCES

[1] 2018. a2c announcement. https://blog.openai.com/baselines-acktr-a2c/. (2018). accessed 09/03/2018.
[2] M. G. Bellemare, Y. Naddaf, J. Veness, and M. Bowling. 2013. The Arcade Learning Environment: An Evaluation Platform for General Agents. *Journal of Artificial Intelligence Research* 47 (jun 2013), 253–279.
[3] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. 2016. OpenAI Gym. (2016). arXiv:arXiv:1606.01540

[4] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. 2672–2680.
[5] Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).
[6] Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, and Brendan Frey. 2015. Adversarial autoencoders. *arXiv preprint arXiv:1511.05644* (2015).
[7] Andrei A Rusu, Neil C Rabinowitz, Guillaume Desjardins, Hubert Soyer, James Kirkpatrick, Koray Kavukcuoglu, Razvan Pascanu, and Raia Hadsell. 2016. Progressive neural networks. *arXiv preprint arXiv:1606.04671* (2016).
[8] Matthew E Taylor and Peter Stone. 2009. Transfer learning for reinforcement learning domains: A survey. *Journal of Machine Learning Research* 10, Jul (2009), 1633–1685.
[9] Jane X Wang, Zeb Kurth-Nelson, Dhruva Tirumala, Hubert Soyer, Joel Z Leibo, Remi Munos, Charles Blundell, Dharshan Kumaran, and Matt Botvinick. 2016. Learning to reinforcement learn. *arXiv preprint arXiv:1611.05763* (2016).
[10] Markus Wulfmeier, Ingmar Posner, and Pieter Abbeel. 2017. Mutual alignment transfer learning. *arXiv preprint arXiv:1707.07907* (2017).
[11] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. 2014. How transferable are features in deep neural networks?. In *Advances in neural information processing systems*. 3320–3328.
[12] Fuzhen Zhuang, Xiaohu Cheng, Ping Luo, Sinno Jialin Pan, and Qing He. 2015. Supervised Representation Learning: Transfer Learning with Deep Autoencoders.. In *IJCAI*. 4119–4125.