Reinforcement Learning in Latent Space

We, the authors

Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Keywords: some important words

1 Introduction

1.1 Research Questions

1.2 Approach

1.3 Related Work

Early work on transfer learning for reinforcement learning mostly relied on human intervention to create a mapping between source and target tasks (e.g. Taylor and Stone, 2007). Taylor and Stone (2007), for example, developed a method called *Rule Transfer*. Their algorithm learns a policy in the source task that gets transformed into rules, serving as advise to the agent when training in the new environment. To use these rules in the target task, hand-coded translation functions were applied. In contrast, Taylor, Kuhlmann, and Stone (2008) published the first system that automatically mapped source and target task. They use little data from a short exploration period in the target task to approximate a one-to-many mapping between the state and action space. This is achieved by comparing all possible state-state and action-action pairs and choosing the ones with the smallest MSE when predicting the next action using neural networks trained on the target task observations. While their method effectively facilitated learning in the target task, it needs to be noted that transfer was performed on modifications of the same task. Hence, they were fairly similar and there was no attempt to tackle cross-domain transfer.

Gupta et al. (2017) used a proxy task learnt in both the source and target domains, and a test task where transfer should occur. Firstly, with the proxy task, pairs of corresponding states are found using time-based alignment or dynamic time warping. Based on these state pairs, a common latent state space is learnt by minimizing reconstruction errors and pairwise distances. In the test task, to incentivize policy transfer from source to task, the distance to source optimal policy in the common space is incorporated the reward function.

Gupta et al. (2018) used the latent space and meta-learning to improve the exploration phase for the new task the agent is tackling. To do so, model-agnostic meta-learning was used to generate knowledge that could be easily adapted for different tasks (latent space). In addition, policy gradients methods in conjunction with meta-learning were utilized to generate and train the policies.

Parisotto, Ba, and Salakhutdinov (2015) trained an agent to learn multiple (related) games of the Atari Learning Environment simultaneously to later generalize from the learned experiences. The training was done by teaching the agent to mimic an expert (hence, actor-mimic) and then doing a feature regression of the learned mimicking. This can be seen as telling the agent what to do and later telling him why he should do it this way. They proposed to use Actor-mimic as a pre-training to

increase learning speed on a set of tasks.

Mnih et al. (2016) used asynchronous gradient descent to train deep neural networks. This framework is lightweight so that it can run on a CPU instead of a GPU. They ”execute multiple agents in parallel on multiple instances of the environment” Mnih et al. (2016). This stabilized learning and reduced the training time. The reduction in training time was roughly linear to the number of processes. Asynchronous advantage actor-critic (A3C) achieved new state-of-the-art performances in 57 Atari games.

Andrychowicz et al. (2017) say that one of the biggest challenges in RL are sparse rewards. They constructed an algorithm that learns from undesired results as well as from desired results. This way the agent can learn from more experiences and thus constructing a reward function is not necessary. Constructing a good reward function is challenging (Ng, Harada, and Russell, 1999) and can be complicated (Popov et al., 2017). They showed that with their approach tasks were able to be learned that previously were not possible. Furthermore, they proposed to train an agent ”on multiple goals even if we care only about one of them.” Andrychowicz et al. (2017).

In the following sections ...

2 Tasks & Experiments

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

3 Results

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

4 Discussion

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis

erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

4.1 Future Work

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

5 Conclusion

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

References

Andrychowicz, Marcin et al. (2017). "Hindsight experience replay". In: *Advances in Neural Information Processing Systems*, pp. 5048–5058.

Gupta, Abhishek et al. (2017). "Learning invariant feature spaces to transfer skills with reinforcement learning". In: *arXiv preprint arXiv:1703.02949*.

Gupta, Abhishek et al. (2018). "Meta-Reinforcement Learning of Structured Exploration Strategies". In: *CoRR* abs/1802.07245. arXiv: 1802.07245. URL: http://arxiv.org/abs/1802.07245.

Mnih, Volodymyr et al. (2016). "Asynchronous methods for deep reinforcement learning". In: *International conference on machine learning*, pp. 1928–1937.

Ng, Andrew Y, Daishi Harada, and Stuart Russell (1999). "Policy invariance under reward transformations: Theory and application to reward shaping". In: *ICML*. Vol. 99, pp. 278–287.

Parisotto, Emilio, Jimmy Lei Ba, and Ruslan Salakhutdinov (2015). "Actor-mimic: Deep multitask and transfer reinforcement learning". In: *arXiv preprint arXiv:1511.06342*.

Popov, Ivaylo et al. (2017). "Data-efficient deep reinforcement learning for dexterous manipulation". In: *arXiv preprint arXiv:1704.03073*.

Taylor, Matthew E., Gregory Kuhlmann, and Peter Stone (2008). "Autonomous Transfer for Reinforcement Learning". In: *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 1*. AAMAS '08. Estoril, Portugal: International Foundation for Autonomous Agents and Multiagent Systems, pp. 283–290. ISBN: 978-0-9817381-0-9. URL: http://dl.acm.org/citation.cfm?id=1402383.1402427.

Taylor, Matthew E. and Peter Stone (2007). "Cross-domain Transfer for Reinforcement Learning". In: *Proceedings of the 24th International Conference on Machine Learning*. ICML '07. Corvalis, Oregon, USA: ACM, pp. 879–886. ISBN: 978-1-59593-793-3. DOI: 10.1145/1273496.1273607. URL: http://doi.acm.org/10.1145/1273496.1273607.