Common Sub-Space Transfer for Reinforcement Learning Tasks[1]

Haitham Bou Ammar [a] Mathew E. Taylor [b] Karl Tuyls [a] Gerhard Weiss [a]

[a] *Department of Knowledge of Engineering, Maastricht University, The Netherlands*
[b] *Department of Computer Science, Lafayette College, PA, USA*

1 Introduction

Agents in reinforcement learning [2] tasks may learn slowly in large or complex tasks– transfer learning [3] is one technique to speed up learning by providing an informative prior. How to best enable transfer between tasks with different state representations and/or actions is currently an open question. This research introduces the concept of a common task subspace, which is used to autonomously learn how two tasks are related. Experiments in two different nonlinear domains empirically show that a learned inter-state mapping can successfully be used by fitted value iteration, to (1) improve the performance of a policy learned with a fixed number of samples, and (2) reduce the time required to converge to a (near-) optimal policy with unlimited samples.

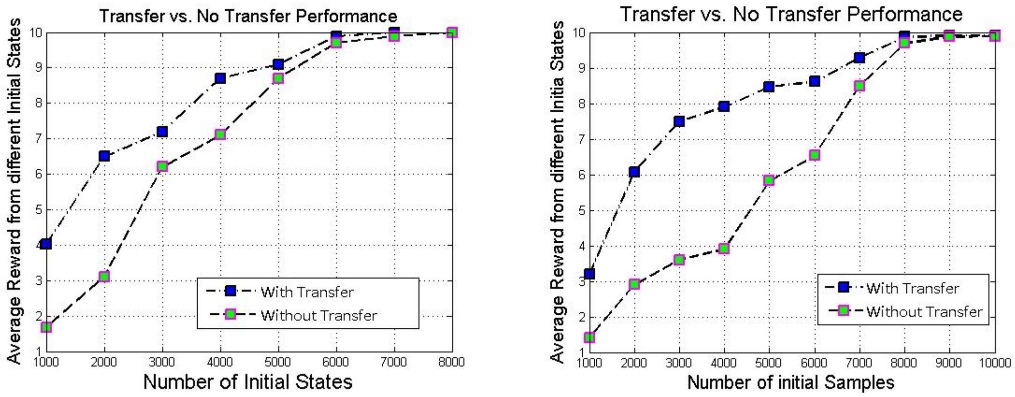
2 Results, Conclusions & Future Work

This research provides a proof-of-concept for our method, using fitted value iteration with locally weighted regression [1] in two experiments. The first experiment shows successful transfer from a single mass system into a double mass system. The second experiment uses a policy learned on the simple inverted pendulum task to improve learning on the cartpole swing-up problem. Although our approach *currently* works in a deterministic model based setting, requires a human specified subspace and is demonstrated using one reinforcement learning algorithm, our results successfully show:

1. an inter-state mapping can be learned from data collected in the source and target tasks with an acceptable number of samples;
2. this inter-state mapping can effectively transfer information from a source task to a target task, even if the state representations and actions differ;
3. an agent that uses transferred information can learn a higher quality policy in the target task, relative to not using this information, when keeping the number of samples in the target task fixed and without using an explicit action mapping; and
4. an agent using information transferred from a source task can learn an optimal policy faster in the target task, relative to not using this information, when it has access to an unlimited number of target task samples thus reducing the number of samples in the target task.

Our approach is composed of three major phases, the first is the determination of the inter-state mapping, relating the state spaces of the tasks, using a common task subspace. It relies on distance measures among state successor state pairs in both task to achieve the goal of finding a correspondence between the state spaces of the two tasks and then conducts a function approximation technique to attain the latter mapping.

[1] This paper was published in the Proceedings of the AAMAS 2011 Workshop on Adaptive Learning Agents and Multi-agent Systems (ALA 2011), Taipei, Taiwan, 2011.



(a) Simple Mass to Double Mass System TransferResults

(b) Inverted Pendulum to Carpole Transfer Results

Figure 1: This figure compares the transfer performance, measured by the averaged reward, vs. different numbers of initial starting states. Starting states can be sampled via the transfer policy or randomly.

While the second is the determination of starting policy in the target task based on similarity transition measures between the two related tasks. This is achieved by mapping state successor states pairs in the target task back to corresponding pairs in the source task and then conducting a search to the most similar transition recommended by the optimal policy of the source task. The action in the target task with the closest similarity to that in the source task accompanied with the intended initial state is used to approximate a good prior in the target task. Lastly, the third is using the thus attained transferred policy as a starting prior for the agent in the target task to improve on.

Our future work will involve three major goals. The first is to extend our algorithms to operate in stochastic model-free MDP settings. The second is to learn a common subspace automatically in both the action and state spaces. Various ideas could be used to achieve such a goal, one of which could be a dimensionality reduction scheme constrained by the common characteristics shared by the different tasks. The third is to test our transfer method with multiple algorithms including policy iteration, Sarsa(λ) and Q-learning.

References

[1] Christopher G. Atkeson, Andrew W. Moore, and Stefan Schaal. Locally weighted learning. *Artif. Intell. Rev.*, 11(1-5):11–73, 1997.

[2] Lucian Busoniu, Robert Babuska, Bart De Schutter, and Damien Ernst. *Reinforcement Learning and Dynamic Programming Using Function Approximators*. CRC Press, Inc., Boca Raton, FL, USA, 1st edition, 2010.

[3] Matthew E. Taylor and Peter Stone. Transfer learning for reinforcement learning domains: A survey. *Journal of Machine Learning Research*, 10:1633–1685, 2009.