Curriculum Learning in Reinforcement Learning

Sanmit Narvekar
Department of Computer Science
University of Texas at Austin
sanmit@cs.utexas.edu

Abstract

Transfer learning in reinforcement learning is an area of research that seeks to speed up or improve learning of a complex target task, by leveraging knowledge from one or more source tasks. This thesis will extend the concept of transfer learning to *curriculum learning*, where the goal is to design a sequence of source tasks for an agent to train on, such that final performance or learning speed is improved. We discuss completed work on this topic, including methods for semi-automatically *generating* source tasks tailored to an agent and the characteristics of a target domain, and automatically *sequencing* such tasks into a curriculum. Finally, we also present ideas for future work.

1 Introduction

As autonomous agents are called upon to perform increasingly difficult tasks, new techniques will be needed to make learning such tasks tractable. Transfer learning [Lazaric, 2011; Taylor and Stone, 2009] is a recent area of research that has been shown to speed up learning on a complex task by transferring knowledge from one or more easier *source tasks*. Most existing transfer learning methods treat this transfer of knowledge as a one-step process, where knowledge from all the sources are directly transferred to the target. However, for complex tasks, it may be more beneficial (and even necessary) to gradually acquire skills over multiple tasks *in sequence*, where each subsequent task requires and builds upon knowledge gained in a previous task.

The goal of this thesis work is to extend transfer learning to the problem of *curriculum learning*. As a motivating example, consider the game of Quick Chess[1] (Figure 1). Quick Chess is a game designed to introduce players to the full game of chess, by using a sequence of progressively more difficult "subgames." For example, the first subgame is a 5x5 board with only pawns, where the player learns how pawns move and about promotions. The second subgame is a small board with pawns and a king, which introduces a new objective: keeping the king alive. In each successive subgame, new elements are introduced (such as new pieces, a larger board, or different configurations) that require learning new skills and building upon knowledge learned in previous games. The final game is the full game of chess.

[1]http://www.intplay.com/uploadedFiles/Game_Rules/P20051-QuickChess-Rules.pdf

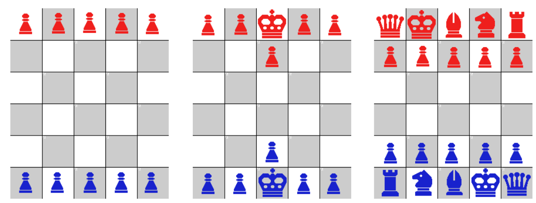

Figure 1: Different subgames in Quick Chess

The question that motivates this line of work is: can we find an optimal sequence of tasks (i.e. a curriculum) for an agent to play that will make it possible to learn some target task (such as chess) fastest, or at a performance level better than learning from scratch?

Designing an effective curriculum is a complex problem that ties task creation, sequencing, and transfer learning. A good set of source tasks is crucial for having positive transfer. A curriculum designer must be able to suggest source tasks using knowledge of the target domain, and that are tailored to the current abilities of the agent. Second, the tasks must be sequenced in a way that allows new knowledge to be accumulated by the agent along each step. Finally, transfer learning must be leveraged to transfer knowledge between tasks in the curriculum. In the following sections, we discuss progress made towards some of these objectives, and plans for future work in this direction.

2 Completed Work

There are two pieces of completed work that relate to this thesis. At AAMAS 2016, we presented the overall problem of curriculum learning in the context of reinforcement learning [Narvekar *et al.*, 2016], and showed concrete multi-stage transfer. Specifically, as a first step towards creating a curriculum, we considered how a space of useful source tasks could be generated, using a parameterized model of the domain and observed trajectories of an agent on the target task. A good source task is one that leverages knowledge about the domain to simplify the problem, and also promotes learning new skills tailored to the abilities of the agent in question. Thus, we proposed a series of functions to semi-automatically create useful, agent-specific source tasks, and showed how they could be used to form components of a curriculum in two challenging multiagent domains. Our results showed that a curriculum can improve learning speed or performance compared to learning on the target task from scratch, even when time spent in source tasks is accounted for (i.e. *strong transfer* [Taylor and Stone, 2009]).

This method is also useful for creating tasks in the classic transfer learning paradigm (1 step curriculum). Past work in transfer learning has typically assumed a fixed set of source tasks are provided, using a static analysis of the domain. This work allowed new source tasks to be suggested from a dynamic analysis of the agent's performance.

While our AAMAS paper proposed a way of creating a space of tasks, it assumed a human expert was available to select and sequence tasks from this space into a curriculum. At IJCAI 2017, we propose a method to address this limitation by automating the sequencing of tasks [Narvekar *et al.*, 2017]. We formalize the generation of a curriculum as a Markov Decision Process, which explicitly models the progress of an agent as it learns through a sequence of tasks. We describe how a policy over such a curriculum MDP can be used to produce a curriculum, and propose an algorithm that approximates one execution of an optimal policy in this MDP. The algorithm is evaluated in a grid world domain to produce curricula tailored to the sensing and action capabilities of 3 different agents. Our results show that the curricula produced improve learning compared to learning without a curriculum, but also that having curricula *customized* for each agent makes a significant difference.

3 Directions for Future Work

Our IJCAI paper shows that different agents can benefit from different curricula. However, it goes through the process of generating a curriculum for each agent independently. This process requires collecting extensive experience in the source tasks for each agent. Thus, generating a curriculum for use by a single agent is generally not practical. An interesting question for future work then is how can we adapt a curriculum generated for one agent to a new agent?

Another interesting direction is characterizing what type of knowledge is useful to extract from a task, and whether we can combine multiple forms of transfer for use in a curriculum. Both of our completed works show how one form of transfer – value function transfer – can be used to transfer knowledge via a curriculum. However, transfer learning literature has shown that alternate forms of transfer are also possible, using options [Soni and Singh, 2006], policies [Fernández *et al.*, 2010], models [Fachantidis *et al.*, 2013], or even samples [Lazaric *et al.*, 2008]. For example, it may be that some tasks are more suitable for transferring options, while others are more suitable for transferring samples.

Finally, our IJCAI paper considers automatically sequences tasks in a single-agent reinforcement learning domain. This is because the formulation of the curriculum MDP is over the policy space of a single agent. However, it would be interesting to extend this idea to a multiagent setting. While we could simply consider a joint policy space, a more efficient solution could be possible using ideas from multiagent systems research.

Acknowledgements

Part of this work was done with Jivko Sinapov, Matteo Leonetti, and Peter Stone as part of the Learning Agents Research Group (LARG) at the University of Texas at Austin. LARG research is supported in part by NSF (CNS-1330072, CNS-1305287, IIS-1637736, IIS-1651089), ONR (21C184-01), AFOSR (FA9550-14-1-0087), Raytheon, Toyota, AT&T, and Lockheed Martin. Peter Stone serves on the Board of Directors of Cogitai, Inc. The terms of this arrangement have been approved by the University of Texas at Austin in accordance with its policy on objectivity in research.

References

[Fachantidis *et al.*, 2013] Anestis Fachantidis, Ioannis Partalas, Grigorios Tsoumakas, and Ioannis Vlahavas. Transferring task models in reinforcement learning agents. *Neurocomputing*, 107:23–32, 2013.

[Fernández *et al.*, 2010] Fernando Fernández, Javier García, and Manuela Veloso. Probabilistic policy reuse for inter-task transfer learning. *Robotics and Autonomous Systems*, 58(7):866 – 871, 2010. Advances in Autonomous Robots for Service and Entertainment.

[Lazaric *et al.*, 2008] A. Lazaric, M. Restelli, and A. Bonarini. Transfer of samples in batch reinforcement learning. In *Proceedings of the Twenty-Fifth Annual International Conference on Machine Learning (ICML-2008)*, Helsinki, Finalnd, July 2008.

[Lazaric, 2011] A. Lazaric. Transfer in reinforcement learning: a framework and a survey. In M. Wiering and M. van Otterlo, editors, *Reinforcement Learning: State of the Art*. Springer, 2011.

[Narvekar *et al.*, 2016] Sanmit Narvekar, Jivko Sinapov, Matteo Leonetti, and Peter Stone. Source task creation for curriculum learning. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016)*, May 2016.

[Narvekar *et al.*, 2017] Sanmit Narvekar, Jivko Sinapov, and Peter Stone. Autonomous task sequencing for customized curriculum design in reinforcement learning. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, August 2017.

[Soni and Singh, 2006] Vishal Soni and Satinder Singh. Using homomorphisms to transfer options across continuous reinforcement learning domains. In *Proceedings of American Association for Artificial Intelligence (AAAI)*, 2006.

[Taylor and Stone, 2009] Matthew E. Taylor and Peter Stone. Transfer learning for reinforcement learning domains: A survey. *Journal of Machine Learning Research*, 10(1):1633–1685, 2009.