Autonomous Inter-Task Transfer in Reinforcement Learning Domains

Matthew E. Taylor
Department of Computer Sciences
The University of Texas at Austin
Austin, Texas 78712-1188
mtaylor@cs.utexas.edu
http://www.cs.utexas.edu/∼mtaylor

Overview

In *reinforcement learning* (RL) (Sutton & Barto 1998) problems, agents take sequential actions with the goal of maximizing a reward signal, which may be time-delayed. In recent years RL tasks have been gaining in popularity as learning methods able to handle complex problems. RL algorithms, unlike many machine learning approaches, do not require correctly labeled training examples and thus may address a wide range of difficult and interesting problems. If RL agents begin their learning *tabula rasa*, mastering tasks may be slow or infeasible. A significant amount of current research in RL thus focuses on improving the speed of learning by exploiting domain expertise with varying degrees of autonomy.

My thesis will examine one such general method for speeding up learning: *transfer learning*. In transfer learning problems, a *source task* can be used to improve performance on, or speed up learning in, a *target task*. An agent may thus leverage experience from an earlier task to learn the current task. A common formulation of this problem presents an agent with a pair of tasks and the agent is told explicitly to train on one before the other. Alternately, in the spirit of *multitask learning* (Caruana 1995) or *lifelong learning* (Thrun 1996), an agent could consult a library of past tasks that it has mastered and transfer knowledge from one or more of them to speed up the current task.

Transfer learning in RL is an important topic to address at this time primarily for three reasons. Firstly, RL techniques have, in recent years, achieved notable successes in difficult tasks which other machine learning techniques are either unable or ill-equipped to address (e.g., TDGammon (Tesauro 1994), elevator control (Crites & Barto 1996), Keepaway (Stone, Sutton, & Kuhlmann 2005), and Server Job Scheduling (Whiteson & Stone 2006)). Secondly, classical machine learning techniques are sufficiently mature that they may now easily be leveraged to assist with transfer learning. Thirdly, promising initial results show that not only are such transfer methods possible, but they can be very effective at speeding up learning.

When physical or virtual agents are deployed, any mechanism that allows for faster learned responses to a new task has the potential to greatly improve their efficacy. Thus, any transfer method that is able to handle the above differences could potentially be utilized by such agents to increase their adaptability and performance when an agent must perform a new task.

With motivations similar to those of *case based reasoning* (Agnar & Enric 1994), where a symbolic learner constructs partial solutions to the current task from past solutions, a primary goal of transfer learning is to autonomously determine how a current task is related to a previously mastered task and then to automatically use past experience to learn faster. My thesis focuses on the following question:

> Given a pair of related RL tasks that have different state spaces, different applicable actions, and/or different representative state variables, how and to what extent can agents transfer knowledge from the source task to learn faster in the target task, and what, if any, domain knowledge must be provided to the agent?

The primary contribution of this thesis will be to address the above question, demonstrating a series of techniques that are able to successfully transfer knowledge between tasks with varying degrees of similarity and given domain knowledge. There are many ways of formulating and addressing the transfer learning problem, but we distinguish this work from previous transfer work (Selfridge, Sutton, & Barto 1985; Singh 1992; Asada *et al.* 1994; Maclin *et al.* 2005; Fernandez & Veloso 2006; Soni & Singh 2006) in three ways:

1. Our methods focus on allowing differences in the action space, the state, and state variables between the two tasks, increasing their applicability relative to many existing transfer methods. However, we will show that they are also applicable when the transition function, reward function, and/or initial state differ.

2. Our methods are competitive with, or are able to outperform, other transfer methods with similar goals.

3. Our methods are able to *learn* relationships between pairs of tasks without relying on human domain knowledge, a necessity for achieving autonomous transfer.

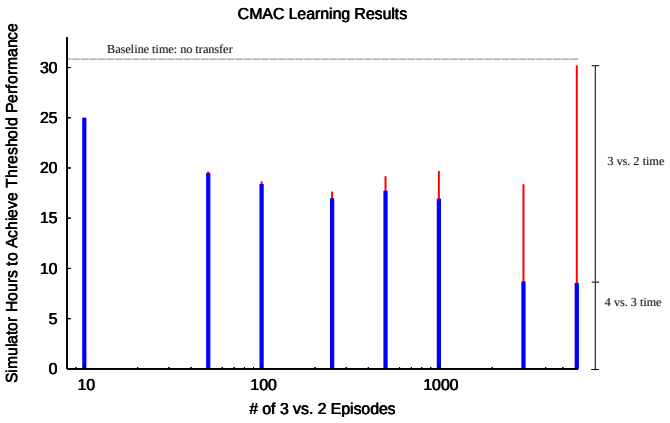

Figure 1: This graph demonstrates that both the target task training time and the total training time can be reduced via value function transfer. The x-axis shows the number of episodes spent learning in the source task (3 vs. 2 Keepaway) and the y-axis shows the amount of time needed to reach the threshold performance in the target task (4 vs. 3 Keepaway).

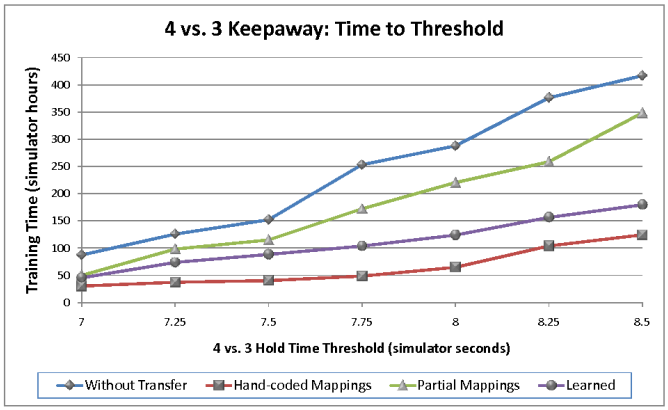

Figure 2: This graph shows the results of policy search transfer between 3 vs. 2 and 4 vs. 3 Keepaway. The x-axis shows the target task threshold performance and the y-axis shows the target task learning time required to reach the threshold. Learning without transfer is compared to learning after transfer with partial inter-task mappings (which utilizes incomplete information about the relationship between the two tasks), transfer with learned inter-task mappings, and transfer with inter-task mappings provided by an oracle.

Core contributions

I now enumerate the components we consider necessary to address the main question posed in this thesis.

1. **Problem Definition:** Our transfer problems will focus on using a *source task* to speed up learning in a *target task* and I will define the scope of such problems in a RL setting.
2. **Performance Metrics:** In order to measure the efficacy of our methods I have defined two transfer-specific metrics. I argue that the two metrics are appropriate for the RL domains considered and focus on the performance speedup due to transfer, rather than the performance of a particular underlying TD or policy search *base learning algorithm*. Both metrics measure the amount of time learners take to reach a threshold performance in the target task. *Target task learning time* measures the amount of time that learners take to reach the threshold performance with and without transfer; time spent in the source task is ignored. Transfer is successful if the target task can be learned faster with transfer than without. *Total learning time* measures the total amount of time spent training. Without transfer, only time in the target task is counted; when using transfer, both the source and target task training time must be accounted for. Transfer is successful by this more difficult metric if it is faster to learn the source and target tasks via transfer than to learn the target task directly.
3. **Oracle-Enabled Transfer:** One class of transfer methods considered utilize inter-task mappings. Inter-task mappings describe relations between state variables and actions in the source and target tasks; they are used so that learned knowledge in the source task can apply to a target task even when the state and action spaces have changed. I first assume that an oracle provides mappings that are complete and correct. We investigate three distinct ways of using the inter-task mappings: for value function transfer (Taylor, Stone, & Liu 2005), for policy search transfer (Taylor, Whiteson, & Stone 2007), and via rule transfer (Taylor & Stone 2007). An example of reducing both the target task training time and total training time via value function transfer can be seen in Figure 1.
4. **Learning Task Relationships:** I also consider pairs of tasks where no oracle exists and the inter-task mapping must be learned. Constructing such relationships is the primary difficulty when transferring between disparate tasks, but I plan to leverage a variety of existing machine learning techniques to assist with this process. I will demonstrate the effectiveness of these relationship-learning methods on pairs of related tasks and then combine them with the above transfer methods to achieve autonomous transfer. Thus far we have succeeded by making the (strong) assumption that objects can be described by a constant set of state variables, regardless of the task these object appear in (Taylor, Whiteson, & Stone 2007). An example of how the target task training time can be reduced in policy search transfer with and without provided inter-task mappings is shown in Figure 2.
5. **Empirical Validation:** To validate our transfer methods, I will fully implement them in at least three domains. Success in different domains and with different implementations, which have different qualitative characteristics, will show that our methods have broad applicability as well as significant impact. Thus far we have concentrated on the Robosoccer Keepaway Domain (Taylor, Stone, & Liu 2005) and the Server Job Scheduling Domain (Taylor, Whiteson, & Stone 2007).

Supplemental Contributions

In addition to these goals, I am considering at least two supplemental goals, but am actively searching for more goals so that my thesis can more fully develop our understanding of transfer:

1. **Inter-Domain Transfer:** I informally define a *domain* to be a setting for a group of semantically similar *tasks*. While many methods exist to transfer between domains, none have been shown to work between domains. In addition to showing that inter-domain transfer is feasible (Taylor & Stone 2007), I would like to show that such transfer can be done autonomously.
2. **Effects of Task Similarity on Transfer Efficacy:** All the RL tasks I consider can be parameterized and thus it is possible to make the source and target tasks more or less similar. For instance, preliminary results in the Keepaway domain show that transfer is able to improve learning, compared to learning without the benefit of transfer, when the players in the two tasks have pass actuators with different accuracies, but transfer is more beneficial when the players in both tasks have actuators with the same capabilities. Observing, and ideally predicting, how transfer degrades as the source and target tasks become more dissimilar should lead to a better understanding of the proposed transfer methods. Such heuristics could be used to determine if two tasks are "similar enough" that transfer could provide any benefit. Defining a similarity metric for tasks based on these heuristics would also potentially allow us to *construct* a source task for a given target task.

Acknowledgements

I would like to thank Peter Stone, my coauthors, and the anonymous reviewers. This research was supported in part by DARPA grant HR0011-04-1-0035, NSF CAREER award IIS-0237699, EIA-0303609.

References

Agnar, A., and Enric, P. 1994. Case-based reasoning: Foundational issues, methodological variations, and system approaches.

Asada, M.; Noda, S.; Tawaratsumida, S.; and Hosoda, K. 1994. Vision-based behavior acquisition for a shooting robot by using a reinforcement learning. In *Proc. of IAPR/IEEE Workshop on Visual Behaviors-1994*, 112–118.

Caruana, R. 1995. Learning many related tasks at the same time with backpropagation. In *Advances in Neural Information Processing Systems 7*, 657–664.

Crites, R. H., and Barto, A. G. 1996. Improving elevator performance using reinforcement learning. In Touretzky, D. S.; Mozer, M. C.; and Hasselmo, M. E., eds., *Advances in Neural Information Processing Systems 8*, 1017–1023. Cambridge, MA: MIT Press.

Fernandez, F., and Veloso, M. 2006. Probabilistic policy reuse in a reinforcement learning agent. In *Proc. of the 5th International Conference on Autonomous Agents and Multiagent Systems*, 720–727.

Maclin, R.; Shavlik, J.; Torrey, L.; Walker, T.; and Wild, E. 2005. Giving advice about preferred actions to reinforcement learners via knowledge-based kernel regression. In *Proceedings of the 20th National Conference on Artificial Intelligence*.

Selfridge, O. G.; Sutton, R. S.; and Barto, A. G. 1985. Training and tracking in robotics. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, 670–672.

Singh, S. P. 1992. Transfer of learning by composing solutions of elemental sequential tasks. *Machine Learning* 8:323–339.

Soni, V., and Singh, S. 2006. Using homomorphisms to transfer options across continuous reinforcement learning domains. In *Proceedings of the Twenty First National Conference on Artificial Intelligence*.

Stone, P.; Sutton, R. S.; and Kuhlmann, G. 2005. Reinforcement learning for RoboCup-soccer keepaway. *Adaptive Behavior* 13(3):165–188.

Sutton, R. S., and Barto, A. G. 1998. *Introduction to Reinforcement Learning*. MIT Press.

Taylor, M. E., and Stone, P. 2007. Cross-domain transfer for reinforcement learning. In *Proceedings of the Twenty-Fourth International Conference on Machine Learning*.

Taylor, M. E.; Stone, P.; and Liu, Y. 2005. Value functions for RL-based behavior transfer: A comparative study. In *Proceedings of the Twentieth National Conference on Artificial Intelligence*.

Taylor, M. E.; Whiteson, S.; and Stone, P. 2007. Transfer via inter-task mappings in policy search reinforcement learning. In *The Sixth International Joint Conference on Autonomous Agents and Multiagent Systems*.

Tesauro, G. 1994. TD-Gammon, a self-teaching backgammon program, achieves master-level play. *Neural Computation* 6(2):215–219.

Thrun, S. 1996. Is learning the n-th thing any easier than learning the first? In Touretzky, D. S.; Mozer, M. C.; and Hasselmo, M. E., eds., *Advances in Neural Information Processing Systems*, volume 8, 640–646. The MIT Press.

Whiteson, S., and Stone, P. 2006. Evolutionary function approximation for reinforcement learning. *Journal of Machine Learning Research* 7(May):877–917.