Categorizing Transfer for Reinforcement Learning

Matthew E. Taylor TAYLORM@USC.EDU
Computer Science Department, University of Southern California, Los Angeles, CA

Peter Stone PSTONE@CS.UTEXAS.EDU
Department of Computer Sciences, University of Texas at Austin, Austin, TX

1. Transfer in Reinforcement Learning

The insight behind *transfer learning* (TL) is that generalization may occur not only within tasks, but also *across tasks*. This insight is not new; transfer has long been studied in the psychological literature, such as Thorndike & Woodworth (1901) and Skinner (1953). More relevant are a number of approaches that transfer between machine learning tasks (c.f., Thrun (1996)). However, TL for Reinforcement Learning (RL) tasks has only recently been gaining attention in the artificial intelligence community. Our proposed poster provides a framework in which to describe TL methods that transfer from one or more RL *source task(s)* to an RL *target task*.

2. A Proposed Taxonomy

The taxonomy we propose consists of a six-dimensional classification of algorithms, enumerated below.

I) Transfer Setting / Metrics for Success: Transfer techniques assume varying degrees of autonomy and make many different assumptions. To be fully autonomous, an RL transfer agent would have to perform all of the following steps:

1. Given a target task, select an appropriate source task or set of tasks from which to transfer.
2. Learn how the source task(s) and target task are related.
3. Effectively transfer knowledge from the source task(s) to the target task.

While the mechanisms used for these steps will necessarily be interdependent, no TL methods are currently capable of robustly accomplishing all three goals. TL methods can be classified according to which question(s) they answer.

II) Task Difference Assumptions: What assumptions does the TL method make about how the source and target are allowed to differ? Examples of things that can differ between the source and target tasks include different system dynamics (i.e., the target task becomes harder to solve is some incremental way), or different sets of possible actions at some states. Such assumptions define the types of source and target tasks that the method can transfer between. Allowing transfer to occur between less similar source and target tasks gives more flexibility to a human designer in the human-guided scenario. In the fully autonomous scenario, more flexible methods are more likely to be able to successfully apply past knowledge to novel target tasks.

There is also a question regarding the distribution from which the source task and target target task come from. First, in a *multi-task learning* setting, both distributions are typically identical, and the agent is expected to learn how to learn quickly over tasks in this distribution. Second, an embodied agent may face many different tasks over the course of its lifetime, and be asked to transfer between them (Sutton et al., 2007). In this case, the tasks are drawn from some distribution based on the agents environment, which may change over time. Third, if a human is selecting the source task(s), they may be from a distribution that is arbitrarily different from the target task.

III) Source Task Selection: In the simplest case, the agent assumes that it should use all source tasks, and that they are from the same distribution as the target task. Another option would be for a human to perform source task selection; the agent must transfer all the source task(s) to the target task, even if they are drawn from different distributions. More complex TL methods allow the agent to select a source task, or set of source tasks, from which to transfer.

A source task selection mechanism may be designed to guard against *negative transfer*, where transfer hurts the learner's performance. The more robust the selection mechanism, the more likely it is that transfer will be able to provide a benefit. While no definitive answer to this problem exists, successful techniques will likely have to account for specific target task characteristics.

IV) Task Mappings: Many methods require a mapping to transfer effectively: in addition to knowing that a source task and target task are related, they need to know *how* they are related. *Inter-task mappings* (Taylor et al., 2007) are a way to define how two tasks are related. If a human is in the loop, the method may assume that such task mappings are provided; if the agent is expected to transfer autonomously, such mappings have to be learned. Different methods use a variety of techniques to enable transfer, both on-line (while learning the target task) and offline (after learning the source task but before learning the target task). Such learning methods attempt to minimize the number of samples needed and/or the computational complexity of the method, while still learning a mapping to enable effective transfer.

V) Type of Knowledge Transferred: What type of information is transferred between the source and target tasks? This information can range from very low-level information about a specific task (i.e., the expected outcome when performing an action in a particular location) to general heuristics that attempt to guide learning. Different types of knowledge may transfer better or worse depending on task similarity. For instance, low-level information may transfer across closely related tasks, while high-level concepts may transfer across pairs of less similar tasks. The mechanism that transfers knowledge from one task to another is closely related to what is being transferred, how the task mappings are defined (if they are indeed necessary), and what assumptions exist about the similarity of the two tasks.

VI) Compatible Learning Methods: Does the TL method place restrictions on what RL algorithm is used, such as applying only to temporal difference methods? Different learning algorithms have different biases. Ideally an experimenter or agent would select the RL algorithm to use based on characteristics of the task, not on the TL algorithm. Some TL methods require that the source and target tasks be learned with the same method, other allow a class of methods to be used in both tasks, but the most flexible methods decouple the agents' learning algorithms in the two tasks.

3. Contributions of Poster

The proposed MSRL poster will serve three purposes:

1. It will expose unfamiliar viewers to transfer learning in RL.
2. The proposed TL classification will be explained in the context of current methods.
3. The poster will suggest future research directions, based on what is missing from existing work, in the context of the proposed taxonomy.

References

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, *4*, 237–285.

Skinner, B. F. (1953). *Science and human behavior*. Colliler-Macmillian.

Sutton, R. S., Koop, A., & Silver, D. (2007). On the role of tracking in stationary environments. *Proceedings of the 24th International Conference on Machine Learning*.

Taylor, M. E., Stone, P., & Liu, Y. (2007). Transfer learning via inter-task mappings for temporal difference learning. *Journal of Machine Learning Research*, *8*, 2125–2167.

Thorndike, E., & Woodworth, R. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review*, *8*, 247–261.

Thrun, S. (1996). Is learning the n-th thing any easier than learning the first? *Advances in Neural Information Processing Systems* (pp. 640–646).

Thrun, S., & Pratt, L. (Eds.). (1998). *Learning to learn*. Norwell, MA, USA: Kluwer Academic Publishers.