Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)

Transfer Learning for Multiagent Reinforcement Learning Systems*.

Felipe Leno da Silva and Anna Helena Reali Costa
Escola Politécnica da Universidade de São Paulo, São Paulo, Brazil
{f.leno,anna.reali}@usp.br

Abstract

Reinforcement learning methods have successfully been applied to build autonomous agents that solve many sequential decision making problems. However, agents need a long time to learn a suitable policy, specially when multiple autonomous agents are in the environment. This research aims to propose a Transfer Learning (TL) framework to accelerate learning by exploiting two knowledge sources: (i) previously learned tasks; and (ii) advising from a more experienced agent. The definition of such framework requires answering several challenging research questions, including: *How to abstract and represent knowledge, in order to allow generalization and posterior reuse?*, *How and when to transfer and receive knowledge in an efficient manner?*, and *How to evaluate the transfer quality in a Multiagent scenario?*.

1 Context and Motivation

Reinforcement Learning (RL) [Sutton and Barto, 1998] is an extensively used technique for autonomous agents with the ability to learn through experimentation. First an action that affects the environment is chosen, then the agent observes how much that action collaborated to the task completion through a reward function. An agent can learn how to optimally solve tasks by executing this procedure multiple times. The main limitation of RL is that agents take a long time to learn how to solve tasks. However, like in human learning, previous knowledge can greatly accelerate the learning of a new task. For example, it is easier to learn Spanish beforehand knowing Portuguese (or a similar language).

Many RL domains can be treated as *Multiagent Systems* (MAS), in which multiple agents are acting in a shared environment. We are specially interested in Cooperative Multiagent RL (MARL), in which all agents work cooperatively to solve the same task. In such domains, other type of knowledge reuse is applicable. Agents can communicate to transfer learned behaviors. In the language learning example, being taught by a fluent speaker of the desired language can accelerate learning, because the teacher can identify learner's mistakes and provide customized explanations and examples. However, learning how to actuate in a MAS may be a difficult task, since the environment becomes non-stationary due to the parallel actuation of multiple agents.

Transfer Learning (TL) [Taylor and Stone, 2009] allows to reuse knowledge acquired in previous tasks, and has been used to accelerate learning in RL domains and alleviate scalability issues. In MARL, TL can either reuse knowledge from previously learned tasks or from agent communication, in which one agent can transfer learned behaviors to another agent. Even though TL has been used in many ways in MARL, there is no consensual answer to many aspects that must be defined in order to specify a TL algorithm. This research aims to specify a TL framework to allow knowledge reuse in multiagent domains from both previously learned tasks (when available) and agent tutoring, two scenarios that are common in human learning.

2 Research Goals and Expected Contributions

This research aims to **propose a Transfer Learning framework** to allow knowledge reuse **in Multiagent Reinforcement Learning**, both from previous tasks and among agents. Specifying such method requires the definition of: (i) A model which allows knowledge generalization; (ii) What information is transferred through tasks or agents; and (iii) How to define when the knowledge of a given agent must be transferred to another. Figure 1 depicts the proposed framework. The agent extracts knowledge from advice given by other agents and previously solved tasks to accelerate the learning of a new task. The solution of this new task can then be abstracted and added to the knowledge base.

3 Background and Related Work

Single-agent RL domains are usually modeled as a *Markov Decision Process* (MDP), which can be solved by RL. An MDP is described by the tuple $\langle S, A, T, R \rangle$ [Puterman, 2005], where S is the set of environment states, A is the set of actions available to an agent, T is the transition function, and R is the reward function, which gives a feedback toward task completion. At each decision step, an agent observes the state s and chooses an action a (among the applicable ones

*This research is supported by CNPq (grant 311608/2014-0) and São Paulo Research Foundation (FAPESP), grant 2015/16310-4

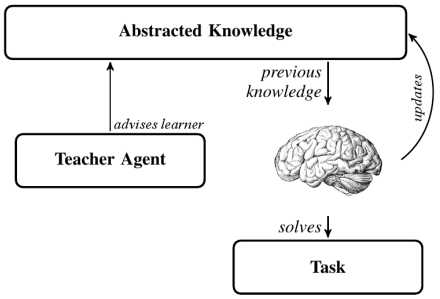

Figure 1: The proposed Transfer Learning framework.

in s). Then the next state is defined by T. The agent must learn a policy π that maps the best action for each possible state. The solution of an MDP is an optimal policy π^*, a function that chooses an action maximizing future rewards at every state. In learning problems (where R and T are unknown) the agent can iteratively learn a Q-table, i.e. a function that maps every combination of state and action to an estimate of the long-term reward starting from the current state-action pair, which eventually converges to the true Q function: $Q^*(s,a) = E\left[\sum_{i=0}^{\infty} \gamma^i r_i\right]$, where γ is a discount rate and r_i is the reward received after i steps from using action a on state s. Q^* can be used to define an optimal policy as: $\pi^*(s) = \arg\max_a Q^*(s,a)$.

However, learning Q^* may take a long time, and TL methods can be used to accelerate convergence. The basic idea in any TL algorithm is to reuse previously acquired knowledge in a new task. In order to use TL in practice, three aspects must be defined [Pan and Yang, 2010]: *What*, *when*, and *how* to transfer. Even though many methods have been developed, there is no consensual definition of how to represent knowledge and how to transfer it. Therefore, the success of a TL method depends on the knowledge representation. The standard MDP formulation is not a good representation for knowledge abstraction, which is of utmost importance when transferring knowledge between two similar tasks.

Relational representations achieved great success in knowledge abstraction and TL [Koga *et al.*, 2015]. *Object-Oriented MDP* (OO-MDP) [Diuk *et al.*, 2008] is a relational representation that allows generalization opportunities by modeling similar entities of a domain as *objects* that follow the description of a *class*. An OO-MDP requires the definition of a set of *classes* C, where each class C_i is composed of a set of *attributes*, and each attribute has a *domain*, which specifies the set of values this attribute can assume. The state is defined by the set of objects that exist in the environment. As the objects of the same class follow the same description, the learner can abstract experiences by assuming that objects of the same class are affected in the same way by actions. For example, a robot learning how to navigate in an environment can learn that moving towards a specific wall results in a disadvantageous collision. Then, the robot can assume that moving towards any wall would be harmful, thus avoiding all wall collisions even though the agent has never collided with many of the walls in the environment.

Although OO-MDP seems promising to TL, so far no OO-MDP extensions to MAS are available in the literature. Thus, the specification of an OO-MDP extension to MAS is a good first step toward a knowledge representation that could lead to successful TL.

4 Partial Results

In order to define a representation which allows knowledge generalization, we propose an OO-MDP extension to MAS, called *Multiagent Object-Oriented MDP* (MOO-MDP). This extension if fully described on an article submitted to ECAI 2016 Main Track, in which an algorithm to solve deterministic cooperative MOO-MDPs is also presented.

MOO-MDP is inspired by the insight that each agent in a MAS can be seen as an object, hence the environment is described by a set of agent and environment objects, in which the former can perform autonomous actions and the latter are unreasoning entities. While MOO-MDP enables state space abstraction by generalizing experiences for all objects of the same class, coordinated behaviors can still be learned since agents can identify other reasoning entities and act according to previous experiences with that same class of agents.

5 Next Steps

MOO-MDP is a promising model which allows knowledge generalization. Now, the next step in our research is to define how to transfer learned knowledge through tasks or agents. Abstract policies have been successfully used in single-agent Transfer Learning [Koga *et al.*, 2015], thus we now plan to build abstract policies based on MOO-MDPs and transfer significant parts of them through similar tasks. We still need to specify a mapping method to find correspondences between states and actions in different domains, and how the transfer of knowledge among agents may be executed with abstract policies.

References

[Diuk *et al.*, 2008] C. Diuk, A. Cohen, and M. L. Littman. An Object-oriented Representation for Efficient Reinforcement Learning. In *Int. Conf. on Machine Learning (ICML)*, pages 240–247, 2008.

[Koga *et al.*, 2015] M. L. Koga, V. F. da Silva, and A. H. R. Costa. Stochastic Abstract Policies: Generalizing Knowledge to Improve Reinforcement Learning. *IEEE Transactions on Cybernetics*, 45(1):77–88, 2015.

[Pan and Yang, 2010] S. J. Pan and Q. Yang. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010.

[Puterman, 2005] M. L. Puterman. *Markov Decision Processes : Discrete Stochastic Dynamic Programming*. J. Wiley & Sons, Hoboken (N. J.), 2005.

[Sutton and Barto, 1998] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, USA, 1st edition, 1998.

[Taylor and Stone, 2009] M. E. Taylor and P. Stone. Transfer Learning for Reinforcement Learning Domains: A Survey. *Journal of Machine Learning Research*, 10:1633–1685, 2009.