Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)

Accelerating Multiagent Reinforcement Learning through Transfer Learning

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 (RL) is a widely used solution for sequential decision-making problems and has been used in many complex domains. However, RL algorithms suffer from scalability issues, especially when multiple agents are acting in a shared environment. This research intends to accelerate learning in multiagent sequential decision-making tasks by reusing previous knowledge, both from past solutions and advising between agents. We intend to contribute a *Transfer Learning framework* focused on Multiagent RL, requiring as few domain-specific hand-coded parameters as possible.

Context and Motivation

Reinforcement Learning (RL) (Littman 2015) has been extensively used to solve sequential decision-making problems with minimal input data. Despite the recent successes in many complex domains, the classical RL approach is not scalable and suffers from the *curse of dimensionality*. Scalability issues are further intensified in Multiagent domains, as the decision maker may need to reason over the other agents in the environment. In a world where more and more devices have computing power, many tasks can be solved by Multiagent Systems (sometimes must). Therefore, scalable and robust techniques are of utmost importance for Multiagent RL algorithms. One approach to accelerate the learning process and scale RL to complex domains is to reuse previous knowledge in the new task. Transfer Learning (TL) methods have been used to profit from previous knowledge in many ways (Taylor and Stone 2009). For example, a robot learning how to play soccer may reuse previously acquired skills, such as walking while carrying a ball. Additionally, a more experienced teammate could advice the agent on when to shoot to maximize the goal chance. Also, the agent may observe its teammates' behavior and imitate strategies and moves to learn faster than when learning from scratch without any previous knowledge or guidance.

In this research we aim at accelerating learning in Multiagent RL Systems by reusing knowledge from multiple sources. Even though we here focus on Multiagent RL Systems, the main ideas of our proposal are applicable in the Multiagent Systems, Reinforcement Learning, and Machine Learning areas in general. TL has already been used in Multiagent RL and presented promising results. However, there are no consensual answers to many aspects that must be defined to specify a suitable TL algorithm. Especially, we want to develop algorithms that require as few domain-specific hand-coded definitions as possible. This is a challenging goal, as it is hard to autonomously generalize knowledge, define task similarities, and reuse task solutions.

Research Goals and Expected Contributions

This research aims to **propose a Transfer Learning framework** to **accelerate Multiagent Reinforcement Learning** through knowledge reuse. Figure 1 depicts the proposed framework. The agent has an *Abstract Knowledge Base* from which it can extract previous solutions of tasks that are similar to the new one. Additionally, an *Advisor Agent* may provide advice helping the agent to achieve its goals. The agent then is expected to learn the task faster than without TL. When the task is solved, the new task solution can be abstracted and stored in the knowledge base for latter use.

Specifying an algorithm that integrates this framework into the learning processes requires the definition of: (i) A model which allows knowledge generalization; (ii) What information is transferred through tasks or agents; (iii) How to identify which previous tasks are similar enough to be used; (iv) How to define when the knowledge of a given agent must be transferred to another; and (v) How to define if an agent giving advice is trustworthy.

Background and Related Work

RL is a popular solution for sequential decision-making tasks modeled as *Markov Decision Processes* (MDP). An MDP is described by the tuple $\langle S, A, T, R \rangle$, where S is the set of environment states, A is the set of available actions, T is the transition function, and R is the reward function, which gives a feedback toward task completion. At each decision step, the agent observes the state s and chooses an action a (among the applicable ones 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

*We gratefully acknowledge financial support from CNPq (grant 311608/2014-0) and São Paulo Research Foundation (FAPESP), grant 2015/16310-4.

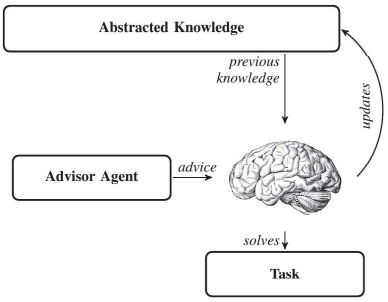

Figure 1: The proposed Transfer Learning framework.

an MDP is an optimal policy π^*, a function that chooses an action maximizing future rewards at every state.

However, in learning problems (where R and T are unknown), learning π^* may take a long time, and TL methods can be used to accelerate convergence. The basic idea of 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: *What*, *when*, and *how* to transfer.

Even though many methods have been developed, there is no consensual definition of how to represent and transfer knowledge. In the *teacher-student* framework (Taylor et al. 2014),a teacher agent observes a student agent and suggests actions for states in which applying the wrong action is expected to achieve bad results in the long term. Although achieving a relevant speed-up in single-agent tasks, the *teacher-student* framework requires a teacher able to perform the task with expertise when the learning starts. That is, if all agents start to learn together they are likely to give bad advice, which can hurt the learning process instead of accelerating it. Also, the student must be observed constantly, which is unfeasible for many Multiagent Systems.

Partial Results and Future Works

An adaptation of the *teacher-student* framework focused on systems composed of simultaneously learning agents will be proposed in a submission to the main track of AAMAS 2017. The agent relations in our proposal are termed *advisor-advisee* relations, where the advisor not necessarily has to perform optimally. Instead of having a fixed teacher, the advisee evaluates its confidence in the current state, and broadcasts an advice request for all the reachable friendly agents in case its confidence is not high enough. Each prospective advisor then evaluates its own confidence in the advisee's state. In case the advisor's confidence is high enough, an *ad hoc advisor-advisee* relation is initiated and the advisor suggests an action. This procedure is more robust than receiving advice from a single agent, since in case one of the advisors suggests a wrong action the correct action might still be identified combining multiple advice. Advice work as a heuristic for the exploration strategy, thus it does not affect the convergence of most base learning algorithms (after the budget is spent the agents return to their standard exploration strategy). We assume in our initial work that the agents are cooperatively solving a task or that all agents participating in advice relations are members of a team. However, it may be of interest of further works to devise trust mechanisms to apply the *ad hoc* advising in MAS composed of agents with different (or unknown) objectives. Assuming that the agent policy becomes better as the states are repeatedly explored, one of the proposed confidence metrics is to derive the confidence from the number of times the agent visited the state before. Our proposal seems to be a promising way to provide the advising ability of Figure 1. So far, we have explored the benefits of the *ad hoc* advising in robot soccer simulations, both when all agents are learning from scratch and when one expert agent is present in the system when the learning starts. In our initial experiments, our proposal presented a speed-up when compared to state-of-the-art advising techniques in both scenarios.

The next long term step in my proposal is to define how to build and reuse the *Abstract Knowledge Base*. The agents must be able to compute task similarities in order to select the most similar task from the knowledge base and reuse its solution. Also, the task solutions need to be abstracted before stored in the knowledge base, since overfitted solutions are not likely to be useful in new tasks. While a common procedure in the literature is to provide hand-coded task mappings, autonomously computing task similarities is hard. In our BRACIS paper, we propose the *Multiagent Object-Oriented MDP* (MOO-MDP) (Silva, Glatt, and Costa 2016), a relational approach in which the state space is described through classes of objects. The main idea is that each entity in the environment (agent or not) belongs to a class of objects, and then the agent can generalize experiences. For example, in the Robot Soccer domain, if the agent passes the ball to a teammate with a high goal opening angle and it scores a goal, the agent then reasons that any teammate with a high goal opening and the ball possession is likely to score a goal. MOO-MDP improved learning in our initial experiments through generalization, and now we need to define a way to compute task similarities and reuse task solutions. Comparing the attributes of the objects in the environment seems to be a promising way to define a task similarity metric, but how to compute this metric is still an open problem.

Finally, the last step in our proposal will be to properly integrate the two types of knowledge reuse consistently.

References

Littman, M. L. 2015. Reinforcement Learning Improves Behaviour from Evaluative Feedback. *Nature* 521(7553):445–451.

Silva, F. L. d.; Glatt, R.; and Costa, A. H. R. 2016. Object-Oriented Reinforcement Learning in Cooperative Multiagent Domains. In *Proceedings of the 5th Brazilian Conference on Intelligent Systems (BRACIS)*, 19–24.

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

Taylor, M. E.; Carboni, N.; Fachantidis, A.; Vlahavas, I. P.; and Torrey, L. 2014. Reinforcement Learning Agents Providing Advice in Complex Video Games. *Connection Science* 26(1):45–63.