Integrating Agent Advice and Previous Task Solutions in Multiagent Reinforcement Learning

Doctoral Consortium

Felipe Leno Da Silva
University of São Paulo, Brazil

ABSTRACT

Reinforcement learning methods have successfully been applied to build autonomous agents that solve challenging sequential decision-making problems. However, agents need a long time to learn a task, especially when multiple autonomous agents are in the environment. This research aims to propose a Transfer Learning framework to accelerate learning by combining two knowledge sources: (i) previously learned tasks; and (ii) advice 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 consistently combine knowledge from several sources?*

KEYWORDS

Transfer Learning; Reinforcement Learning; Multiagent Learning

ACM Reference Format:
Felipe Leno Da Silva. 2019. Integrating Agent Advice and Previous Task Solutions in Multiagent Reinforcement Learning. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019,* IFAAMAS, 2 pages.

1 CONTEXT AND MOTIVATION

Reinforcement Learning (RL) is an extensively used technique to train autonomous agents 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, but RL agents require a huge number of interactions to learn. However, like in human learning, reuse of previous knowledge can greatly accelerate the learning process. 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. In such domains, another 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 mistakes and provide customized explanations and examples.

Transfer Learning (TL) [4] allows to reuse previously acquired knowledge, and has been used to accelerate learning in RL domains and alleviate scalability issues. In Multiagent RL (MARL), TL methods have been applied to reuse both internal knowledge from previously learned tasks and learned behaviors from agent communication separately, but no work combined them. This research aims to specify a TL framework to allow knowledge reuse by combining both previously learned task solutions and agent advice, two scenarios that are common in human learning.

Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 13–17, 2019, Montreal, Canada.

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. In order to specify a method to fulfill the expected contributions, we need to define: (i) A model which allows knowledge generalization; (ii) What information is transferred through tasks or agents; (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 ($\mathscr{K}^{agents}$) and combines it with previously solved tasks ($\mathscr{K}^{source}$) to accelerate the learning of a new task. The solution of this new task ($\mathscr{K}^{target}$) can then be abstracted and added to the knowledge base. In the long-term, the agent is expected to learn tasks much faster due to the task solutions stored in his knowledge base and the received advice, which is specific for the current task.

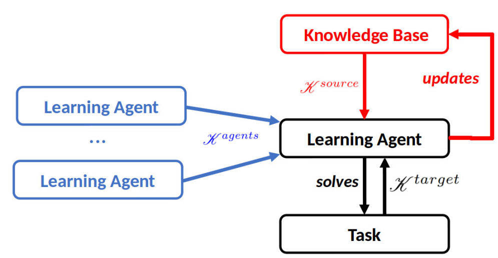

Figure 1: The proposed Transfer Learning framework.

Even though we here focus on MARL, the main ideas of our proposal are applicable in the Multiagent Systems, Reinforcement Learning, and Machine Learning areas in general.

3 BACKGROUND AND RELATED WORK

Single-agent sequential decision problems are often modeled as a *Markov Decision Process* (MDP), which can be solved by RL. An MDP is described by the tuple $\langle S, A, T, R \rangle$, 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 towards task completion. At each decision step, an 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 an MDP is an optimal policy π^*, a function that chooses an action maximizing future rewards at every state. In learning problems the agent usually estimates the quality of each action through exploring the state-action space and observing the received reward signal. However, learning this estimate may take a long time, and TL methods can be used to accelerate learning. The basic idea in any TL algorithm is to reuse acquired knowledge.

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 knowledge and how to transfer it.

In the *teacher-student* framework [11], a more experienced agent (teacher) suggests actions to a learning agent (student). However, works following the *teacher-student* paradigm assume that teachers follow a fixed (and good) policy. This means that, in order to apply this idea in a Multiagent RL domain, teacher-student relations could only be established after teachers have trained enough to achieve a fixed policy, but we are concerned about systems composed of simultaneously learning agents, where this assumption does not hold. For the reuse of knowledge from previous tasks, varied types of information have been successfully transferred, such as samples of low-level interactions with the environment [9], policies, value functions [10], abstract or partial policies, and heuristics or biases for a more effective exploration, each of them presenting benefits over learning from scratch [7]. As no work combined advice-based procedures with those methods, we are not sure yet on which information would be optimal for transfer in our setting.

4 PARTIAL RESULTS AND FUTURE WORK

Our first step towards the framework described in Section 2 was the development of an advising framework based on *teacher-student*, called *Ad Hoc Advising* [5], that is specialized to tasks in which multiple agents are learning together.

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 agents in case its confidence is low. Each prospective advisor then evaluates its own confidence in the advisee's state. In case the advisor's confidence is high, an ad hoc advisor-advisee relation is initiated and the advisor suggests an action. Advice works as a heuristic for the exploration strategy, thus it does not affect the convergence of most base learning algorithms (after the maximum number of advice is spent the agents return to their standard exploration strategy). 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 and our proposal presented a speed-up when compared to state-of-the-art advising techniques.

We have also carried out some exploratory works on how to abstract and transfer knowledge across tasks (without combining it with agent advise yet). So far we have mostly been exploring the generalization provided by *object-oriented* representations [6]. Our first work leveraging this representation estimates Probabilistic Inter-TAsk Mappings (PITAM) [2] through human-given task descriptions. The main idea is to receive a relational description of each task and a class mapping to relate entities in the two tasks. Based on that, the algorithm estimates a probabilistic mapping from one task to another, which can be used to TL. The *object-oriented* representation has also been used to decompose complex tasks into smaller ones, that are faster to solve and from which knowledge can be reused to learn the complex task faster [3].

Our next step for this Ph.D. research is to figure out how to bridge the two types of TL and combine those knowledge reuse techniques. We are currently exploring the use of Distributional RL [1] to use the learned distribution as a proxy of the confidence of the agent in the current policy, which could possibly be used as a confidence function for *ad hoc advising* after reusing knowledge from previous tasks. Many other lines remain open for exploration. An especially prominent one is the security aspect of transfer procedures. How can the agent be robust against malicious communications? An argumentation or trust mechanism to evaluate the advice quality would be needed. Most of the transfer algorithms in the literature also require previously-defined communication protocols for transfer of information. Therefore, it would be interesting to develop methods for the *Ad Hoc Teamwork* [8] setting, where the other agents in the system are previously unknown and no commonly-known protocol is available at the beginning of the learning process.

ACKNOWLEDGMENTS

We acknowledge financial support from CNPq, grants 425860/2016-7 and 307027/2017-1 and São Paulo Research Foundation (FAPESP), grants 2015/16310-4, 2016/21047-3, and 2018/00344-5.

REFERENCES

[1] Marc G. Bellemare, Will Dabney, and Rémi Munos. 2017. A Distributional Perspective on Reinforcement Learning. *CoRR* abs/1707.06887 (2017).

[2] Felipe Leno Da Silva and Anna Helena Reali Costa. 2017. Towards Zero-Shot Autonomous Inter-Task Mapping through Object-Oriented Task Description. In *Workshop on Transfer in Reinforcement Learning (TiRL)*.

[3] Felipe Leno Da Silva and Anna Helena Reali Costa. 2018. Object-Oriented Curriculum Generation for Reinforcement Learning. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. 1026–1034.

[4] Felipe Leno Da Silva and Anna Helena Reali Costa. 2019. A Survey on Transfer Learning for Multiagent Reinforcement Learning Systems. *Journal of Artificial Intelligence Research (JAIR)* 69 (2019), 645–703.

[5] Felipe Leno Da Silva, Ruben Glatt, and Anna Helena Reali Costa. 2017. Simultaneously Learning and Advising in Multiagent Reinforcement Learning. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. 1100–1108.

[6] Felipe Leno Da Silva, Ruben Glatt, and Anna H. R. Costa. 2019. MOO-MDP: An Object-Oriented Representation for Cooperative Multiagent Reinforcement Learning. *IEEE Transactions on Cybernetics* 49, 2 (2019), 567–579.

[7] Felipe Leno Da Silva, Matthew E. Taylor, and Anna Helena Reali Costa. 2018. Autonomously Reusing Knowledge in Multiagent Reinforcement Learning. In *International Joint Conference on Artificial Intelligence (IJCAI)*. 5487–5493.

[8] Peter Stone, Gal A. Kaminka, Sarit Kraus, and Jeffrey S. Rosenschein. 2010. Ad Hoc Autonomous Agent Teams: Collaboration without Pre-Coordination. In *AAAI Conference on Artificial Intelligence (AAAI)*. 1504–1509.

[9] Ming Tan. 1993. Multi-agent Reinforcement Learning: Independent vs. Cooperative Agents. In *International Conference on Machine Learning (ICML)*. 330–337.

[10] Adam Taylor, Ivana Dusparic, Edgar Galvan-Lopez, Siobhan Clarke, and Vinny Cahill. 2014. Accelerating Learning in Multi-Objective Systems through Transfer Learning. In *International Joint Conference on Neural Networks*. 2298–2305.

[11] Lisa Torrey and Matthew E. Taylor. 2013. Teaching on a Budget: Agents Advising Agents in Reinforcement Learning. In *International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*. 1053–1060.