Methods and Algorithms for Knowledge Reuse in Multiagent Reinforcement Learning

Felipe Leno da Silva, Anna Helena Reali Costa

[1]Escola Politécnica – Universidade de São Paulo (USP)
São Paulo – SP – Brazil

{f.leno,anna.reali}@usp.br

Abstract. *Reinforcement Learning (RL) is a powerful tool that has been used to solve increasingly complex tasks. RL operates through repeated interactions of the learning agent with the environment, via trial and error. However, this learning process is extremely slow, requiring many interactions. In this thesis, we leverage previous knowledge so as to accelerate learning in multiagent RL problems. We propose knowledge reuse both from previous tasks and from other agents. Several flexible methods are introduced so that each of these two types of knowledge reuse is possible. This thesis adds important steps towards more flexible and broadly applicable multiagent transfer learning methods.*

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) [Taylor and Stone 2009] 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 aimed at specifying flexible TL frameworks to allow knowledge reuse by combining both previously learned task solutions and agent advice, individually or in combination, two scenarios that are common in human learning.

2. Research Goals

This research aimed 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 (K^{agents}) and combines it with previously solved tasks (K^{source}) to accelerate the learning of a new task. The solution of this new task (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 its knowledge base and the received advice, which is specific for the current task.

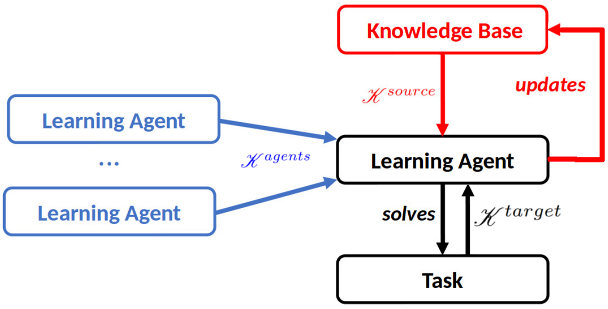

Figure 1. The proposed Transfer Learning framework.

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 [Torrey and Taylor 2013], 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 [Tan 1993], policies, value functions [Taylor et al. 2014], abstract or partial policies, and heuristics or biases for a more effective exploration, each of them presenting benefits over learning from scratch [Silva et al. 2018].

4. Methods and Avenues for 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* [Silva et al. 2017], 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 does not necessarily need 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). We have explored multiple possibilities for designing more efficient confidence functions, including the use of Distributional RL [Bellemare et al. 2017], and methods for estimating the epistemic uncertainty in Deep Learning models [Silva et al. 2020b].

Our proposal was a promising way to provide the advising ability of Figure 1. 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 explored the generalization capabilities provided by *object-oriented* representations [Silva et al. 2019b]. Our first work leveraging this representation estimates Probabilistic Inter-TAsk Mappings (PITAM) [Silva and Costa 2017] 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 [Silva and Costa 2018].

This dissertation opened several avenues of possible research. 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 transfer algorithms in the literature also require communication protocols previously defined for transferring information. Therefore, it would be interesting to develop methods for the *Ad Hoc Teamwork* [Stone et al. 2010] 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.

5. Scientific Results

The work during this Ph.D. was published in several high-impact venues that include the *Journal of Artificial Intelligence Research* (JAIR), *IEEE Transactions on Cybernetics*, *IEEE Transactions on Smart Grid*, *Autonomous Agents and Multi-Agent Systems*, the *AAAI Conference on Artificial Intelligence*, the *International Joint Conference on Artificial Intelligence* (IJCAI), and the *International Conference on Autonomous Agents and Multiagent Systems* (AAMAS).

The following full papers are the main related publications: [Silva et al. 2020a, Silva et al. 2020b, Silva and Costa 2019, Silva et al. 2019b, Silva et al. 2020c, Silva et al. 2018, Silva and Costa 2018, Silva et al. 2017, Silva et al. 2019a, Silva et al. 2016, Silva and Costa 2017].

A *Best Paper Award* from the BRACIS conference was awarded to one of those publications [Silva et al. 2016]. Our work also received a *Honorable Mention as Best Student Poster* at the AAAI Conference on Artificial Intelligence in 2017.

As of submission time, the first author has over 300 citations in total, most of them from the papers above.

6. Main Advances in the State of the Art

In the context of TL, the main objective of this dissertation was to propose methods specialized to reuse knowledge in multiagent RL systems. The high-level idea that a learning agent could reuse knowledge from two sources – previously solved tasks and other agents – guided the initial steps of this work. When the candidate started his Ph.D., a number of methods existed solving portions of those problems individually. However, each work had its own set of assumptions (usually very restricting) that were very hard to integrate.

One of the contributions of this thesis was to write two surveys: one clearly organizing the literature on knowledge reuse in multiagent RL [Silva and Costa 2019], discussing the assumptions of each group of papers, and analyzing the difficulty in developing an integrated framework given the current state of the art. The second one [Silva et al. 2020a] focused on the transfer of knowledge between agents, discussing all the current challenges in translating the knowledge from one agent to another.

Another major contribution was the development of the *Ad Hoc Advising* framework [Silva et al. 2017], focused on TL between agents. The main idea of the method is that agents maintain confidence estimates in their policies and, when confidence in their performance is low, they ask other agents for help. If those agents have high confidence, they might answer with an action suggestion, with the intention of improving the learning speed of the system overall. The *Ad Hoc Advising* is based on the widely-known *Teacher-Student* framework [Torrey and Taylor 2013]. However, a key novelty was introduced by the method. With *Ad hoc Advising* all the agents in the system might assume both roles of *Advisor* or *Advisee* according to their confidence, while in all previous teacher-student frameworks the teacher was fixed and assumed to have (nearly) perfect actuation. This was a key assumption, and the potential of having multiple agents simultaneously learning and sharing knowledge in a multiagent system was recognized by other groups who are already working on extensions of the this work following different points of view [Omidshafiei et al. 2019].

This dissertation has also resulted in an object-oriented task description specialized for multiagent RL settings [Silva et al. 2019b]. This description is easy to understand and to specify even to people without expertise in AI, and helps the learning agent to abstract knowledge (important for building TL approaches). In addition to being easier to specify than other similar relational descriptions, the candidate proposed methods to reuse knowledge across tasks [Silva and Costa 2017], and to autonomously decompose hard tasks into smaller ones to facilitate learning [Silva and Costa 2018]. The proposed method might be a first step towards the inclusion of laypeople into task specification and knowledge reuse of RL systems integrated into the real world, and can be combined with *Ad Hoc Advising* to integrate the framework that motivated this research.

Both groups of methods were validated in robot soccer simulations, a challenging benchmark were the observations and actions are continuous. Our implementation leveraged macro-actions composed of several low-level actions for training the RL algorithms, and directly used the continuous observations. The methods proposed in the thesis are far more flexible than the ones available before, and a number of groups are starting to work on this challenge, clearly influenced by our ideas.

7. Societal impact

The methods here proposed have the potential of having strong societal impact as soon as AI systems become pervasive in our daily life. In a near future, AI devices (e.g. robots) manufactured by different companies will have to coordinate to solve tasks without knowledge of each other's inner workings. Our methods could provide the ways for them to coordinate and share knowledge to solve new tasks, and to create repositories of knowledge that could be accessed by any device in a cloud server.

Moreover, our work on object-oriented task descriptions can facilitate the participation of laypeople, specifying tasks to be performed by their personal devices in a natural manner, without requiring technical knowledge from the users.

Acknowledgments

We gratefully acknowledge funding from the São Paulo Research Foundation (FAPESP), grants 2015/16310-4, 2016/21047-3, and 2018/00344-5; *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES), Finance Code 001; and *Fundação para o Desenvolvimento Tecnológico da Engenharia* (FDTE).

References

Bellemare, M. G., Dabney, W., and Munos, R. (2017). A distributional perspective on reinforcement learning. *CoRR*, abs/1707.06887.

Omidshafiei, S., Kim, D., Liu, M., Tesauro, G., Riemer, M., Amato, C., Campbell, M., and How, J. P. (2019). Learning to Teach in Cooperative Multiagent Reinforcement Learning. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*.

Silva, F. L. D. and Costa, A. H. R. (2017). Towards Zero-Shot Autonomous Inter-Task Mapping through Object-Oriented Task Description. In *Workshop on Transfer in Reinforcement Learning (TiRL)*.

Silva, F. L. D. and Costa, A. H. R. (2018). Object-Oriented Curriculum Generation for Reinforcement Learning. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1026–1034.

Silva, F. L. D. and Costa, A. H. R. (2019). A Survey on Transfer Learning for Multiagent Reinforcement Learning Systems. *Journal of Artificial Intelligence Research (JAIR)*, 69:645–703.

Silva, F. L. D., Costa, A. H. R., and Stone, P. (2019a). Building Self-Play Curricula Online by Playing with Expert Agents. In *Proceedings of the 8th Brazilian Conference on Intelligent Systems (BRACIS)*.

Silva, F. L. D. et al. (2020a). Agents Teaching Agents: A Survey on Inter-agent Transfer Learning. *Autonomous Agents and Multi-Agent Systems*, 34(1):9.

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)*, pages 19–24.

Silva, F. L. D., Glatt, R., and Costa, A. H. R. (2017). Simultaneously Learning and Advising in Multiagent Reinforcement Learning. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1100–1108.

Silva, F. L. D., Glatt, R., and Costa, A. H. R. (2019b). MOO-MDP: An Object-Oriented Representation for Cooperative Multiagent Reinforcement Learning. *IEEE Transactions on Cybernetics*, 49(2):567–579.

Silva, F. L. D., Hernandez-Leal, P., Kartal, B., and Taylor, M. E. (2020b). Uncertainty-Aware Action Advising for Deep Reinforcement Learning Agents. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*.

Silva, F. L. D., Nishida, C. E. H., Roijers, D. M., and Costa, A. H. R. (2020c). Coordination of Electric Vehicle Charging through Multiagent Reinforcement Learning. *IEEE Transactions on Smart Grid*, (accepted).

Silva, F. L. D., Taylor, M. E., and Costa, A. H. R. (2018). Autonomously Reusing Knowledge in Multiagent Reinforcement Learning. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 5487–5493.

Stone, P., Kaminka, G. A., Kraus, S., and Rosenschein, J. S. (2010). Ad Hoc Autonomous Agent Teams: Collaboration without Pre-Coordination. In *AAAI Conference on Artificial Intelligence (AAAI)*, pages 1504–1509.

Tan, M. (1993). Multi-agent Reinforcement Learning: Independent vs. Cooperative Agents. In *International Conference on Machine Learning (ICML)*, pages 330–337.

Taylor, A., Dusparic, I., Galvan-Lopez, E., Clarke, S., and Cahill, V. (2014). Accelerating Learning in Multi-Objective Systems through Transfer Learning. In *International Joint Conference on Neural Networks*, pages 2298–2305.

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

Torrey, L. and Taylor, M. E. (2013). Teaching on a Budget: Agents Advising Agents in Reinforcement Learning. In *International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*, pages 1053–1060.