Agents Teaching Agents: A Survey on Inter-agent Transfer Learning

JAAMAS Track

Felipe Leno Silva[1], Garrett Warnell[2], Anna Helena Reali Costa[3], and Peter Stone[4]
[1]Advanced Institute for AI, São Paulo, SP, Brazil [2]Army Research Laboratory, Austin, TX, USA
[3]University of São Paulo, São Paulo, SP, Brazil [4]The University of Texas at Austin, Austin, TX, USA
f.leno@usp.br,garrett.a.warnell.civ@mail.mil,anna.reali@usp.br,pstone@cs.utexas.edu

ABSTRACT

While reinforcement learning (RL) has helped artificial agents solve challenging tasks, high sample complexity is still a major concern. Inter-agent teaching – endowing agents with the ability to respond to instructions from others – has been responsible for many developments towards scaling up RL. RL agents that can leverage instructions can learn tasks significantly faster than agents that cannot take advantage of such instruction. That said, the inter-agent teaching paradigm presents many new challenges due to, among other factors, differences between the agents involved in the teaching interaction. This paper is a summary of our JAAMAS article [15], where we propose two frameworks that provide a comprehensive view of the challenges associated with inter-agent teaching. We highlight state-of-the-art solutions, open problems, prospective applications, and argue that new research in this area should be developed in the context of the proposed frameworks.

ACM Reference Format:
Felipe Leno Silva[1], Garrett Warnell[2], Anna Helena Reali Costa[3], and Peter Stone[4]. 2020. Agents Teaching Agents: A Survey on Inter-agent Transfer Learning. In *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), Auckland, New Zealand, May 9–13, 2020,* IFAAMAS, 3 pages.

1 INTRODUCTION

Autonomous learning in sequential decision making tasks requires the ability to reason over time-delayed feedback while taking into account environmental, sensory, and actuation stochasticity [7]. Although reinforcement learning (RL) methods [2] enable learning under such conditions, off-the-shelf RL methods can suffer from high sample complexity, which limits their effectiveness in complex domains. Leveraging the experience of another, more competent agent [14], i.e., *inter-agent teaching*, has been a successful approache to addressing sample complexity concerns in RL . Although the literature reports successful inter-agent teaching strategies in terms of learning speed, real-world applications present several additional challenges such as differences between the sensors, actuators, and internal representations of the agents involved. We here summarize our article in the JAAMAS [15], where we formulate two inter-agent teaching frameworks: one in which the teacher is responsible for observing the student behavior and initiating the instruction when it is most needed (i.e., *teacher-driven*), and one in which the learner is proactive to ask for instructions when desired (i.e, *learner-driven*). We present a novel and comprehensive organization and description of all steps involved in those frameworks.

Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), B. An, N. Yorke-Smith, A. El Fallah Seghrouchni, G. Sukthankar (eds.), May 9–13, 2020, Auckland, New Zealand.

2 PROBLEM STATEMENT

An inter-agent teaching relationship requires at least two agents, where a *teacher* agent communicates information to a *learner* – presumably with the intention to accelerate learning (hereafter called *instruction*). We define an instruction as any information communicated by a teacher to a learner with the intention of accelerating learning that *(a)* is specialized to the task at hand, *(b)* can be interpreted and assimilated by the learner, *(c)* is made available during training, and *(d)* is devised without detailed knowledge of the learner's internal representations and parameters. Examples of instructions under this definition are *demonstrations* [11], *action advice* [20], and *scalar feedback* [5] on the current learner policy. We assume that teachers are competent in the learner's task, though they need not be more competent than the learner at all times. This paradigm is situated in the overarching area of transfer learning [12], which can be divided into two main subareas [14]: the subarea covered by our article, i.e., *agents teaching agents* (ATA)—where knowledge is transferred across agents, and *single-agent transfer* (SA)—where knowledge from source tasks is reused by the same agent. In ATA, the knowledge is generated with respect to target task, but it belongs to another agent. In SA, on the other hand, the learning agent itself generates the knowledge to be reused, but with respect to different tasks than the one at hand.

3 BACKGROUND

Sequential decision making problems are often modeled as *Markov decision processes* (MDPs) [10]. An MDP is composed of a tuple $\langle S, A, T, R \rangle$, where S is a set of possible states, A is a set of actions that can be executed by the agent, T is a state transition function , and R is a reward function. In these situations, agents may utilize *reinforcement learning* (RL) [16] techniques to learn behaviors through interacting with their environment and observing samples of those functions. Those samples are the only feedback the agent has to learn policies that achieve good task performance. The main challenge of applying RL is that learning can require a large amount of experience. One way to improve RL is by receiving instructions from another, more-experienced agent. This can help a learning agent build good initial policies, disambiguate knowledge, and/or reduce the amount of experience required in order to learn an acceptable policy [12]. Designing a framework that allows for agents to instruct one another, i.e., *inter-agent teaching*, requires

Table 1: List of all inter-agent teaching modules.

Behavior Generation	**Instruction Definition**
Query Definition	- Instruction construction [12]
- Query timing [3, 4, 6, 8, 13]	- Interfacing
- Teacher selection	**Knowledge Update**
- Query construction	- Receiving instruction
Utility Evaluation	- Instruction reliability
- Behavior observation [1, 9]	- Knowledge merging [12]
- Instruction timing [8, 9, 13]	

dealing with a number of challenges. The literature primarily focuses on only a portion of those problems—none has outlined and discussed completely the modules that must be combined to design an efficient and effective framework.

4 PROPOSED FRAMEWORKS

Broadly speaking, we categorize ATA techniques into one of the two following frameworks: *learner-driven* or *teacher-driven*. In the former, the *learner* is responsible for initiating the interaction between agents. Under this framework, the learning agent must first *generate a behavior*, i.e., attempt to perform its task using some initial policy. Then, throughout the course of the learning process, it is up to the learning agent to decide when and how to *define a query* to send to a (potential) teacher. Assuming the query is successfully received, the teaching agent then *evaluates the utility* of actually providing instruction to the learner in the context of the current situation. If the teacher deems the situation worthy of instruction, the teacher then *defines the instruction* and communicates that instruction to the learner. Finally, the learner then *updates its knowledge* in response to the instruction, after which it is ready to initiate another interaction with the teacher and/or resume learning through its own means. In contrast to the learner-driven framework for ATA, in the *teacher-driven framework*, the *teacher* initiates the interaction between agents. The main difference between this framework and the learner-driven framework is that, in this configuration, there is no explicit query generated by the learner. This lack of query means that it is up to the teacher to decide when the instruction takes place. Table 1 summarizes the modules composing each of the frameworks, summarized below. **Behavior Generation –** To start, before any learning can take place, the learner must first generate an initial behavior from which it can start exploring. Generally, RL agents use a random policy, though perhaps better initial policies can also be found using the agent's own experiences from similar previous tasks [19] (i.e., SA). **Query Definition –** In the context of learner-driven approaches, the agent must define *when* (*query timing*), *to whom* (*teacher selection*), and *how* (*query construction*) to ask for instruction. In principle, the agent could receive instructions at every time step [17]. In many applications, though, communication is limited. In general, it is desirable for inter-agent teaching systems to limit the number of queries to only those that are most needed. After the learner determines when to query a prospective teacher, it might have to reason about *whom* to query. Most inter-agent teaching methods assume that the teacher is known and has agreed to provide instructions. Adaptive teacher-definition algorithms have not been the subject of extensive research, and how to automatically identify, engage with, and estimate the trustworthiness of a new teacher is an open area of research. After that, the task of *constructing* the query may have to be considered. This is itself a challenging research problem, involving both adhering to a given query protocol, and deciding what information should be transmitted as part of the query. **Utility Evaluation –** An important component of the interaction is the strategy used to decide if instructions should be provided to the learner, i.e., *utility evaluation*. For teacher-driven approaches, deciding when to *observe* the learner's behavior is important (*behavior observation*). While many methods assume that the teacher will observe the learner during the entire training process [9, 18, 20], constant observation is impractical in many situations. A second fundamental concern in utility evaluation is deciding when to send the instruction (*instruction timing*). One possible solution is to endow the learner with the ability to modify its behavior to indicate when instructions are most needed, e.g., slowing down its actuation when the confidence in its policy is low [9]. **Instruction definition –** After determining that the current state is appropriate for giving an instruction, the question now is how to define and represent the instruction to be transferred. This is especially challenging if agents have different or unknown representations, or different sensors and actuators, requiring some kind of interface or translation to enable communicating the instruction successfully. The first challenge, *instruction construction*, consists of defining how the instruction is encoded. The *interface* by which the teacher communicates its instructions is also an important factor to consider in all teaching frameworks. Such interfaces consist of two critical components: *(a)* the way in which observations of the learner are presented to the teacher, and *(b)* the way in which instructions are presented to the learner. **Knowledge update –** Finally, after the teacher issues an instruction, the learner is faced with the problem of updating its own knowledge using the information contained in that instruction. Major components of this problem include *receiving* the instruction, determining the *reliability* of the instruction, and *merging* the instruction with the learner's existing knowledge.

5 CONCLUSION

Inter-agent teaching methods have played an important role in augmenting RL methods to increase task learning speed. However, existing literature presents solutions for only some of the many challenges involved in designing these inter-agent methods. We here summarize our article [15], where we provided a comprehensive view of these challenges, and also outlined two broad categories in which to organize them, i.e., *learner-driven* and *teacher-driven* methods. We have also discussed the state-of-the-art options available to implement various modules required in these frameworks.

ACKNOWLEDGMENTS

NSF (CPS-1739964, IIS-1724157, NRI-1925082), ONR (N00014-18-2243), FLI (RFP2-000), ARL, DARPA, Lockheed Martin, GM, Bosch, CNPq (425860/2016-7, 307027/2017-1), and FAPESP (2015/16310-4, 2018/00344-5). P. Stone serves as the Executive Director of Sony AI America and receives financial compensation for this work. The terms of this arrangement have been reviewed and approved by the UT Austin in accordance with its policy on objectivity in research.

REFERENCES

[1] Ofra Amir, Ece Kamar, Andrey Kolobov, and Barbara Grosz. 2016. Interactive Teaching Strategies for Agent Training. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*. 804–811.

[2] Ana L. C. Bazzan. 2014. Beyond Reinforcement Learning and Local View in Multiagent Systems. *Künstliche Intelligenz* 28, 3 (2014), 179–189. https://doi.org/10.1007/s13218-014-0312-5

[3] Sonia Chernova and Manuela Veloso. 2009. Interactive Policy Learning through Confidence-Based Autonomy. *Journal of Artificial Intelligence Research (JAIR)* 34, 1 (2009), 1–25.

[4] Kshitij Judah, Alan P Fern, Thomas G Dietterich, and Prasad Tadepalii. 2014. Active Imitation Learning: Formal and Practical Reductions to I.I.D. Learning. *Journal of Machine Learning Research (JMLR)* 15, 1 (2014), 3925–3963.

[5] W. Bradley Knox and Peter Stone. 2009. Interactively Shaping Agents via Human Reinforcement: The TAMER Framework. In *Proceedings of the 5th International Conference on Knowledge Capture*. 9–16.

[6] Guangliang Li, Hayley Hung, Shimon Whiteson, and W Bradley Knox. 2013. Using Informative Behavior to Increase Engagement in the TAMER Framework. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. 909–916.

[7] Michael L. Littman. 2015. Reinforcement Learning Improves Behaviour from Evaluative Feedback. *Nature* 521, 7553 (2015), 445–451. https://doi.org/10.1038/nature14540

[8] Shayegan Omidshafiei, Dong-Ki Kim, Miao Liu, Gerald Tesauro, Matthew Riemer, Christopher Amato, Murray Campbell, and Jonathan P. How. 2019. Learning to Teach in Cooperative Multiagent Reinforcement Learning. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI)*.

[9] Bei Peng, James MacGlashan, Robert Loftin, Michael L Littman, David L Roberts, and Matthew E Taylor. 2016. A Need for Speed: Adapting Agent Action Speed to Improve Task Learning from Non-Expert Humans. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. 957–965.

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

[11] Stefan Schaal. 1997. Learning from Demonstration. In *Advances in Neural Information Processing Systems (NIPS)*. 1040–1046.

[12] 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.

[13] Felipe Leno Da Silva, Ruben Glatt, and Anna Helena Reali Costa. 2017. Simultaneously Learning and Advising in Multiagent Reinforcement Learning. In *Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. 1100–1108.

[14] Felipe Leno Da Silva, Matthew E. Taylor, and Anna Helena Reali Costa. 2018. Autonomously Reusing Knowledge in Multiagent Reinforcement Learning. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*. 5487–5493.

[15] Felipe Leno Da Silva, Garrett Warnell, Anna Helena Reali Costa, and Peter Stone. 2020. Agents Teaching Agents: A Survey on Inter-agent Transfer Learning. *Autonomous Agents and Multi-Agent Systems* 34, 1 (2020), 9.

[16] Richard S. Sutton and Andrew G. Barto. 1998. *Reinforcement Learning: An Introduction* (1st ed.). MIT Press, Cambridge, MA, USA.

[17] Ming Tan. 1993. Multi-agent Reinforcement Learning: Independent vs. Cooperative Agents. In *Proceedings of the 10th International Conference on Machine Learning (ICML)*. 330–337.

[18] Matthew E. Taylor, Nicholas Carboni, Anestis Fachantidis, Ioannis P. Vlahavas, and Lisa Torrey. 2014. Reinforcement Learning Agents Providing Advice in Complex Video Games. *Connection Science* 26, 1 (2014), 45–63. https://doi.org/10.1080/09540091.2014.885279

[19] Matthew E. Taylor and Peter Stone. 2009. Transfer Learning for Reinforcement Learning Domains: A Survey. *Journal of Machine Learning Research (JMLR)* 10 (2009), 1633–1685. https://doi.org/10.1145/1577069.1755839

[20] Lisa Torrey and Matthew E. Taylor. 2013. Teaching on a Budget: Agents Advising Agents in Reinforcement Learning. In *Proceedings of 12th the International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. 1053–1060.