Robotics Reading Group
@ Instituto Superior Técnico

Session #9
27-05-2020

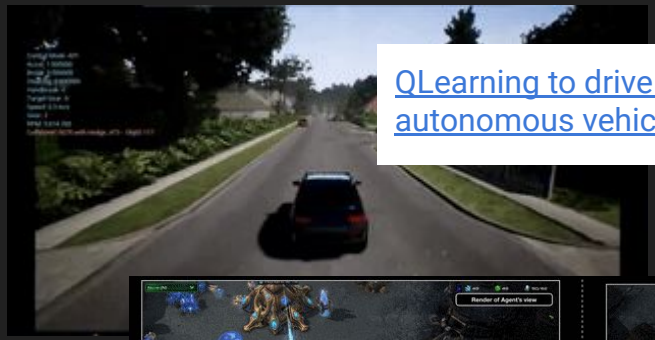
Silvia Tulli

Agents teaching agents: a survey on inter-agent transfer learning

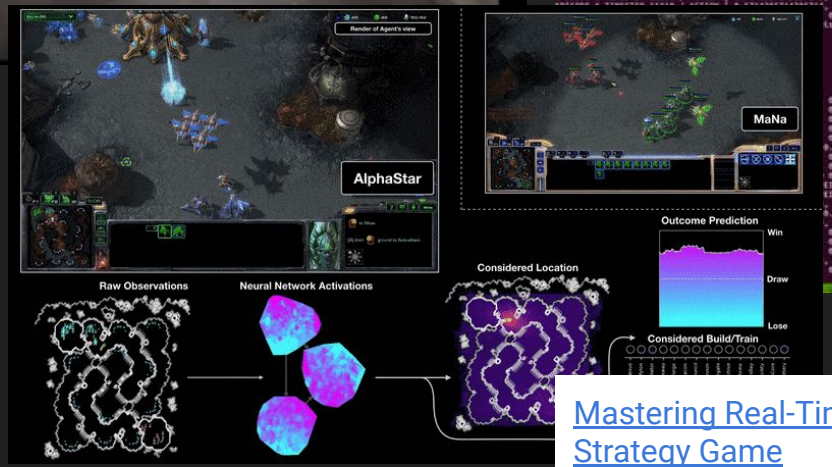
Felipe Leno Da Silva,
Garrett Warnell,
Anna Helena Reali Costa &
Peter Stone

Link to AAMAS2020

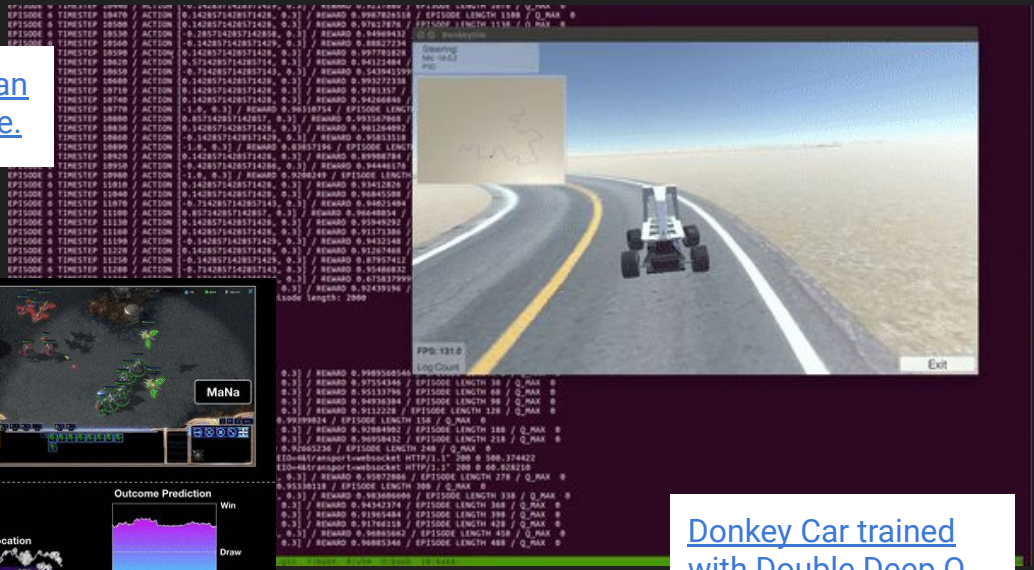
Reinforcement Learning



QLearning to drive an autonomous vehicle.



Mastering Real-Time Strategy Game StartCraft II

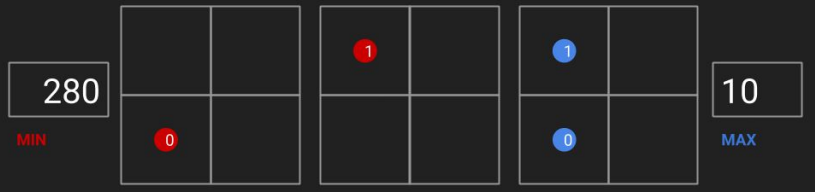


Donkey Car trained with Double Deep Q Learning (DDQN) in Unity Simulator.

Reinforcement Learning

High Sample Complexity

- ~ 2700 episodes for solving Minicomputer Tug of War
- One million of episodes for solving Pommerman



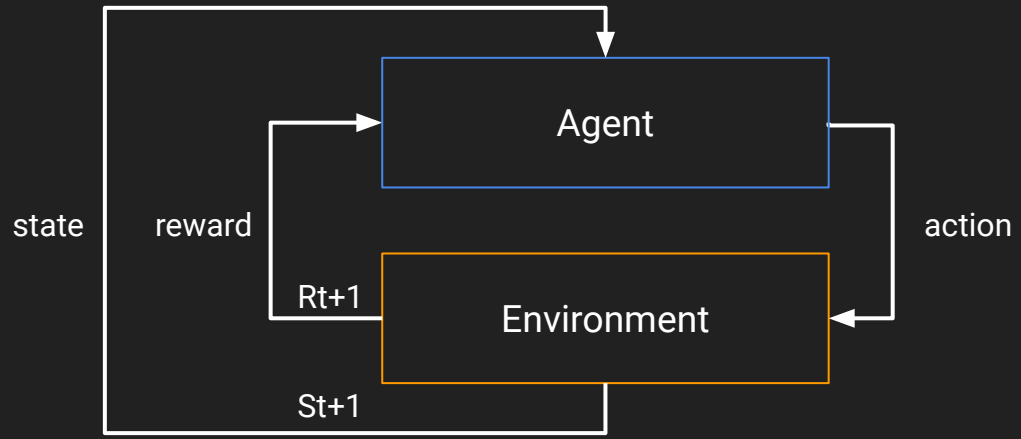
[PlayGround: AI Research into Multi-Agent Learning]

Challenge

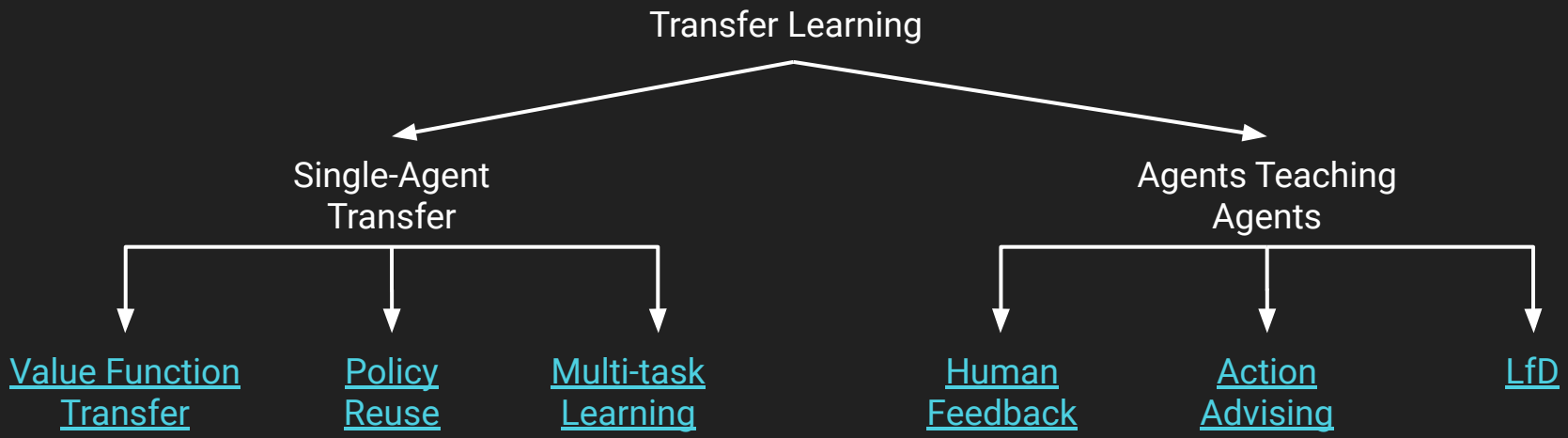
Reinforcement Learning needs ways to accelerate learning

- Why not reuse the experience of another agent?

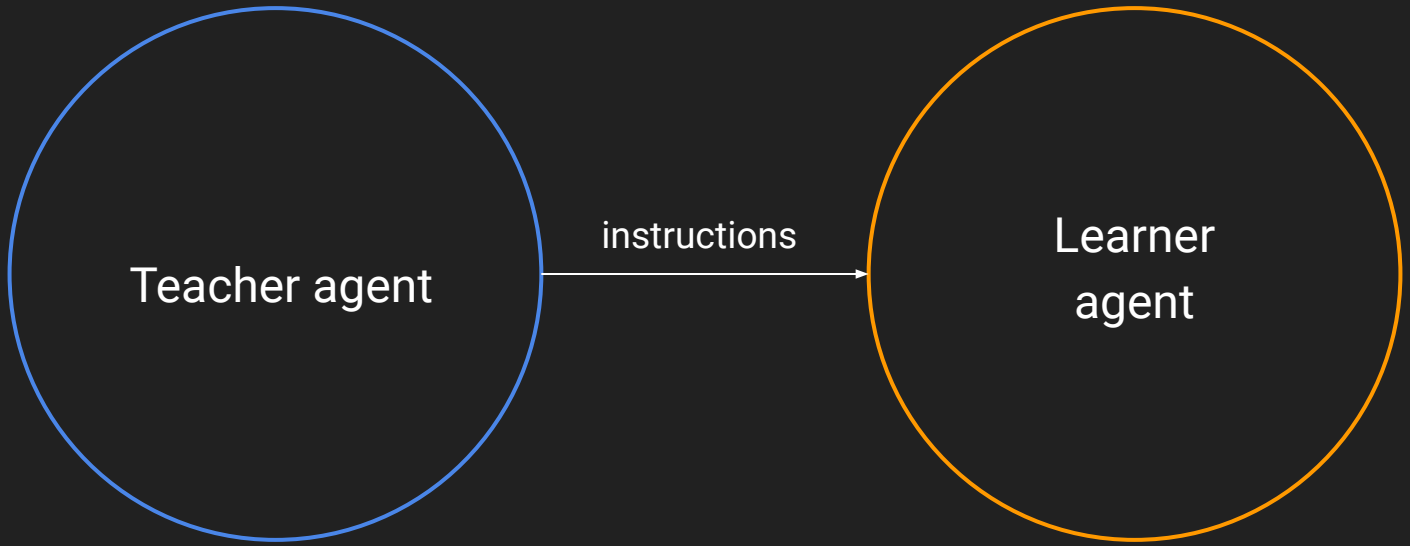
Background



Problem Statement and Scope



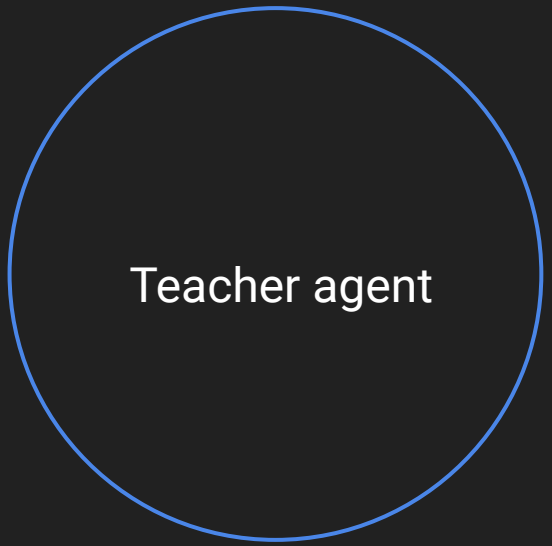
Problem Statement and Scope



Problem Statement and Scope

- Reinforcement Learning agent

Problem Statement and Scope



- RL agent, automated agent following a different algorithm, or human
- it may or may not be learning
- It is competent in the learner's task, though it does not need to be more competent than the learner at all times

Problem Statement and Scope

Instructions are:

- Information specialized to the task at hand
- Information interpreted and assimilated by the learner
- Information made available during training
- Information devised without detailed knowledge of the learner’s internal representations and parameters

Problem Statement and Scope

Instructions examples:

- Demonstrations, action advice, scalar feedback

Instructions are not:

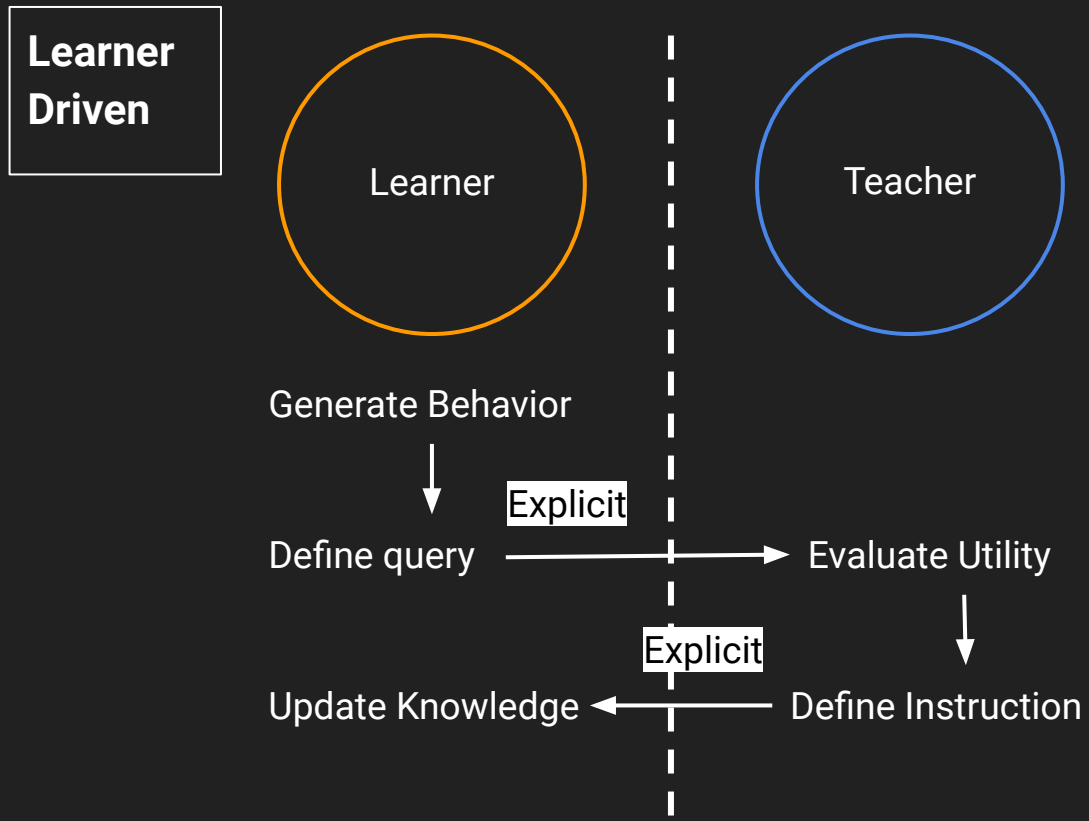
- A reward shaping or a heuristic function built and made available before learning

State of the Art Solutions

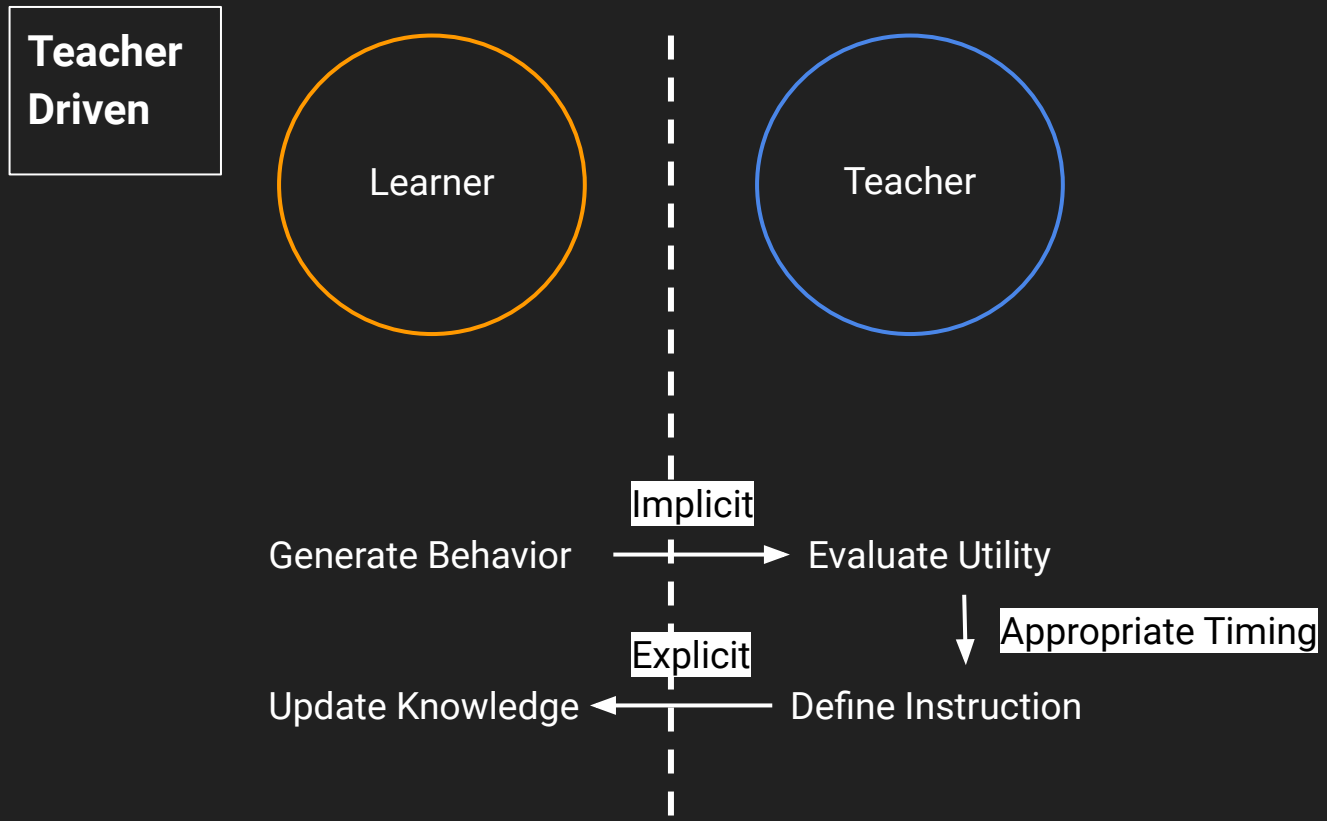
Teacher-Driven

Learner-Driven

Proposed Framework

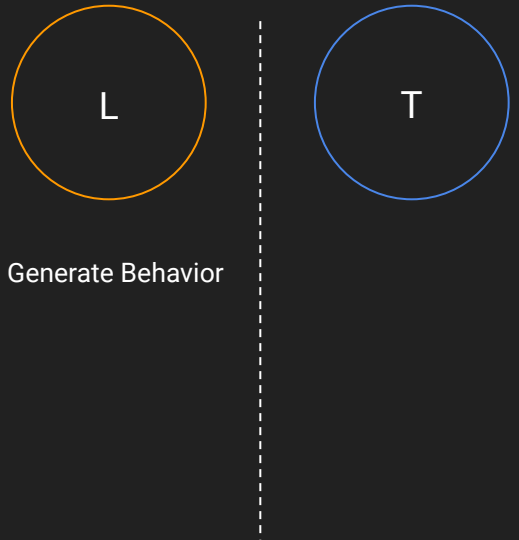


Proposed Framework



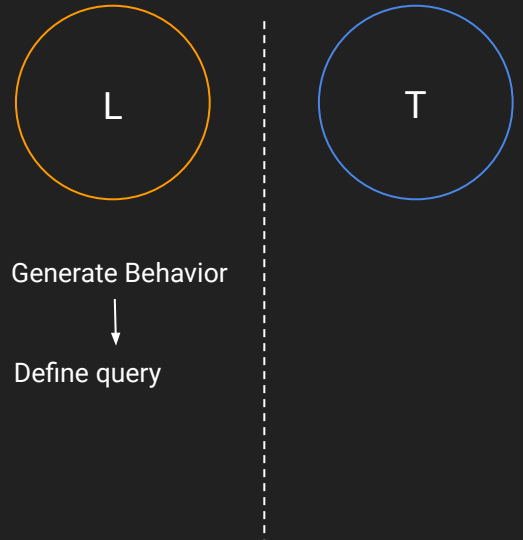
Behavior generation

- Random initialization
- Reusing previous knowledge (single agent transfer)



Query Definition

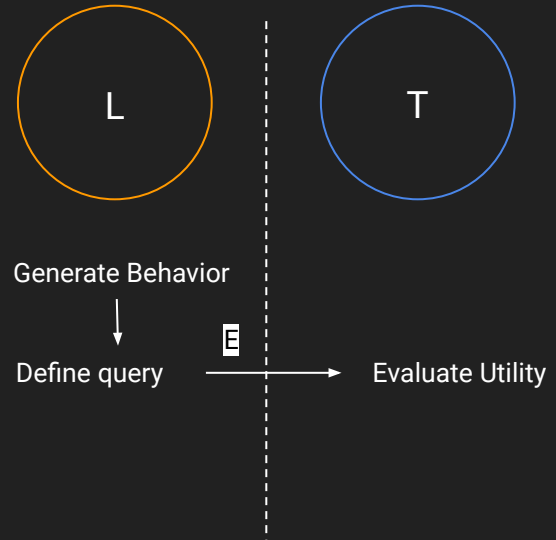
- Query timing (When?)
- Teacher selection (To Whom?)
- Query construction (How? What?)



Utility evaluation

Behavior observation - When to observe the learner's behavior? (in Teacher-driven approaches)

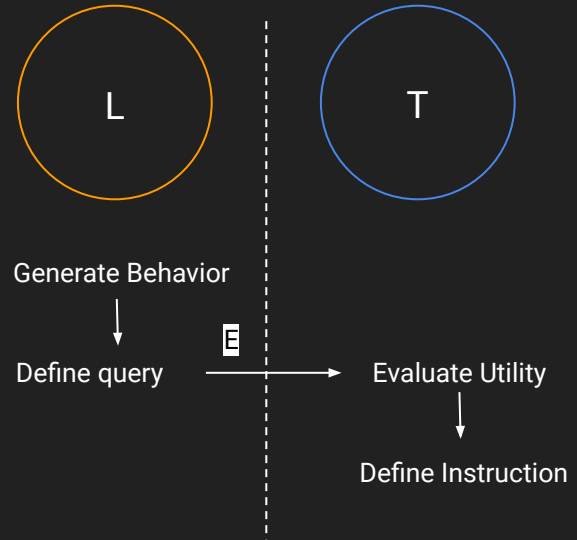
Instruction timing - When to send instruction? What kind of instructions? (in both teacher and learner driven approaches)



Instruction definition

Instruction construction

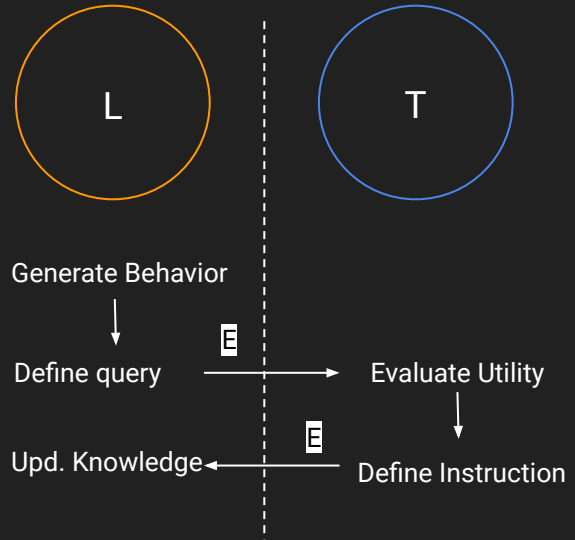
- Action advice (e.g. rules)
- Demonstration
- Natural language instruction
- Preferences
- Feedback (i.e. scalar values, preferences)



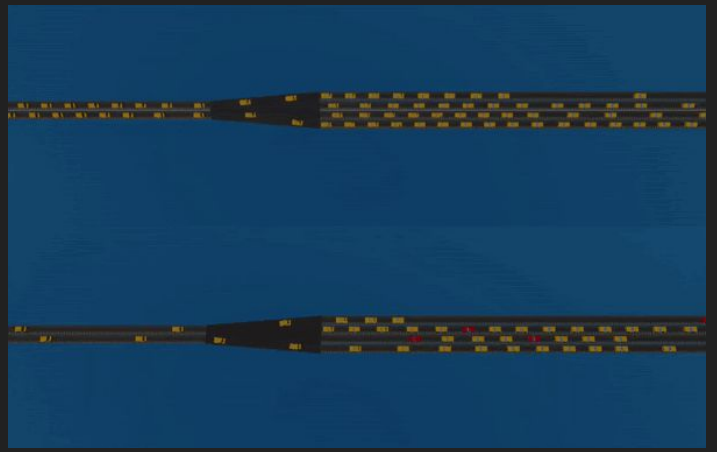
Interfacing and translating instruction

Knowledge Update

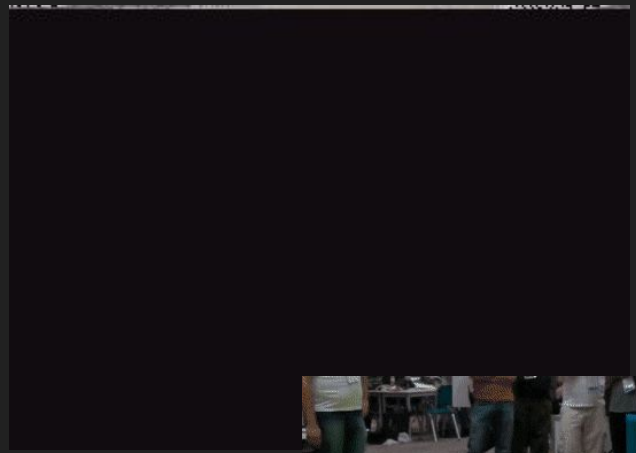
- Receiving instruction:
  - Is the instruction always available?
- Instruction reliability
  - Should the instruction be trusted?
- Knowledge merging
  - Depends by the type of instruction
  - Might be used for guiding exploration



Application examples



Application examples



Challenge

inter-agent teaching

design appropriate inter-agent communication protocols, workable interfaces that consider differences in agent sensors and actuators, and reasonable strategies for translating knowledge if agents use different internal representations

Open Problems

- How to simultaneously use multiple instruction types
- How to adapt instructions to correct undesired behaviors, possibly placing importance on the *behavior observation* challenge
- **Instruction Reliability -** How to explicitly consider that the query or the instruction might be corrupted or lost due to, e.g., a faulty communication channel

Other Discussion Points

- Mutual Modeling
  - How *behavior observation* lead to behavior summarization?
- Curriculum Learning
  - Teacher-Guided Curriculum Learning
- Adhoc Teamwork and Multi-task learning
- Multimodal Transfer Learning
  - Might cross-modality policy transfer be a way to explore how to simultaneously use multiple instruction types?
- Explainable AI and Transfer Learning
- Reinforcement Learning Informed by Natural Language - Human Learning

References

Github repositories:

- Autonomous Driving Cookbook - Distributed Deep Reinforcement Learning for Autonomous Driving
- PlayGround: AI Research into Multi-Agent Learning
- Learning transferable cooperative behaviors in multi-agent teams

Online Articles:

- AlphaStar: Mastering the Real-Time Strategy Game StarCraft II - DeepMind Research
- Intro to Game AI and Reinforcement Learning - Beginner tutorial offered on kaggle
- Curriculum for Reinforcement Learning - great overview written by Lilian Weng, Roboticist for OpenAI - she has a lot of interesting articles on her blog, check it out!

References

Research Papers:

- Crespo, João & Wichert, Andreas. (2020). Reinforcement learning applied to games. SN Applied Sciences. 2. 10.1007/s42452-020-2560-3.
- Narvekar, S., Peng, B., Leonetti, M., Sinapov, J., Taylor, M.E., & Stone, P. (2020). Curriculum Learning for Reinforcement Learning Domains: A Framework and Survey. ArXiv, abs/2003.04960.
- Fournier, P., Sigaud, O., Colas, C., & Chetouani, M. (2019). CLIC: Curriculum Learning and Imitation for feature Control in non-rewarding environments. ArXiv, abs/1901.09720.
- Bengio, Y., Louradour, J., Collobert, R., & Weston, J. (2009). Curriculum learning. ICML '09.
- Silva, F.L., Taylor, M.E., & Costa, A.H. (2018). Autonomously Reusing Knowledge in Multiagent Reinforcement Learning. *IJCAI*.
- Jacq, A., Geist, M., Paiva, A. & Pietquin, O. (2019). Learning from a Learner. Proceedings of the 36th International Conference on Machine Learning, in PMLR 97:2990-2999

References

Research Papers:

- A. Tabrez, S. Agrawal and B. Hayes. (2019) Explanation-Based Reward Coaching to Improve Human Performance via Reinforcement Learning 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Daegu, Korea (South), 2019, pp. 249-257, doi: 10.1109/HRI.2019.8673104.
- Fernando Fernández and Manuela Veloso. (2006). Probabilistic policy reuse in a reinforcement learning agent. In Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems (AAMAS '06). Association for Computing Machinery, New York, NY, USA, 720–727. DOI:https://doi.org/10.1145/1160633.1160762
- Silva, R., Vasco, M., Melo, F.S., Paiva, A., & Veloso, M.M. (2019). Playing Games in the Dark: An approach for cross-modality transfer in reinforcement learning. ArXiv, abs/1911.12851.
- Griffith, S., Subramanian, K., Scholz, J., Isbell, C.L., & Thomaz, A.L. (2013). Policy Shaping: Integrating Human Feedback with Reinforcement Learning. NIPS.

References

Research Papers:

- Silva, Felipe & Hernandez-Leal, Pablo & Kartal, Bilal & Taylor, Matthew. (2020). Uncertainty-Aware Action Advising for Deep Reinforcement Learning Agents.
- Gao, Y., Xu, H., Lin, J., Yu, F., Levine, S., & Darrell, T. (2018). Reinforcement Learning from Imperfect Demonstrations. ArXiv, abs/1802.05313.
- S. V. Albrecht and P. Stone, Autonomous agents modelling other agents: A comprehensive survey and open problems, Artificial Intelligence, vol. 258, pp. 66–95, may 2018, doi: 10.1016/j.artint.2018.01.002.

Questions and Discussion

- What does demonstrations mean in this framework?
    - Demonstrations is considered an instruction type
- Why inter-agent transfer learning reduce the sample size?
    - By accessing to the knowledge of another agent, the learner might reduce the time the agent has to spend in collecting new samples
- How does the paper account for a setting with multiple teacher agents?
- How do we select the teacher?
    - With respect to Q-learning if the protocol is to communicate the state then perhaps the Teacher with the best expected return for that state should be the one answering the query no?
        - We might select by choosing the teacher that perform better - but we need to have access to the other agents behaviors.
- How can we address for the fact that the teacher agent might provide wrong instruction?

Questions and Discussion

- Why can't the learner also inform the teacher to change its internal beliefs and learn from the environment?
  - We could address that using policy extraction for the learner's behavior in order to update the teacher's beliefs, not only about the learner itself, but also about the environment.
- What is the learner perceiving? When will it direct a query to a teacher? When is the teacher confident enough to be assertive?
    - The learner can either decode the encoded information that the teacher shares or receive additional scalar feedback from the teacher. The learner can start by querying the teacher depending on its confidence level for example. The problem of the teacher level of uncertainty needs to be addressed.

Thank you all for coming!