ICML | 2019

Actor-Attention-Critic for Multi-Agent Reinforcement Learning

Shariq Iqbal and Fei Sha

USC University of Southern California

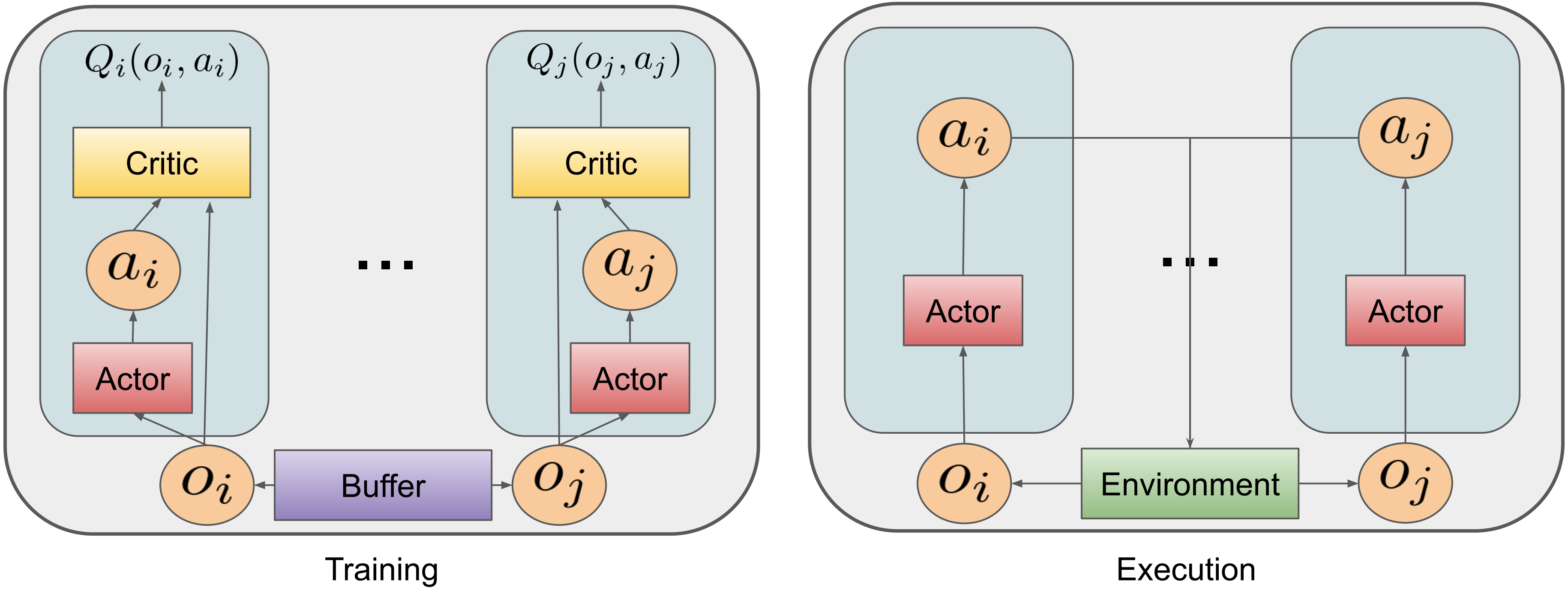
Google AI

Outline

- Establish a baseline approach to MARL
- Demonstrate how recent approaches improve on said baseline through sharing information between agents during training
- Present our attention-based approach for information sharing
- Demonstrate our approach’s improved effectiveness in terms of scalability and overall performance

Baseline Approach to MARL

Learning with single-agent RL technique (actor-critic) for each agent independently

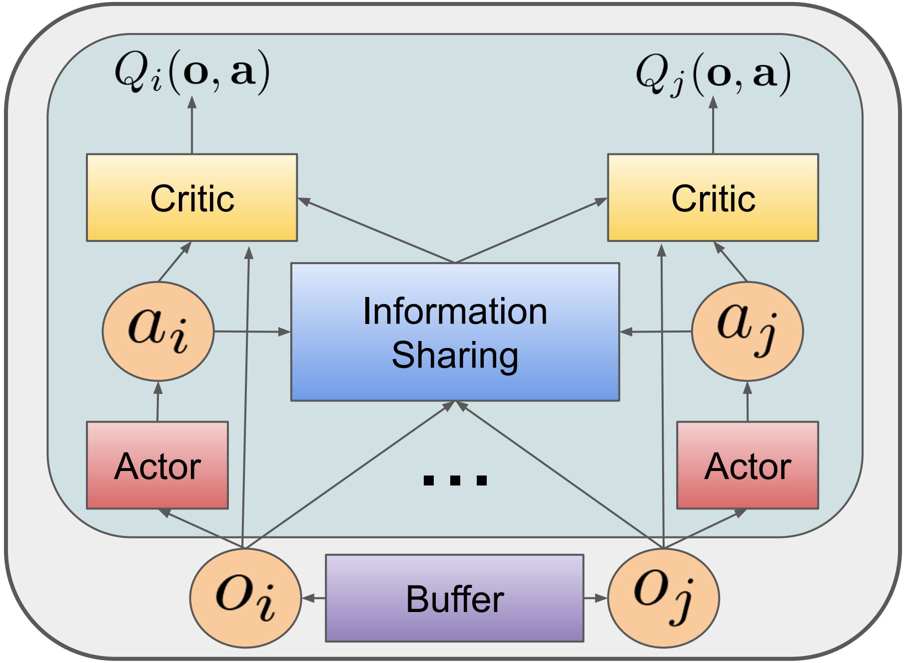


Each agent only considers its local information

Both the actor during execution, and the actor and critic during training

Centralizing Training

Addressing the downsides of the independent MARL approach



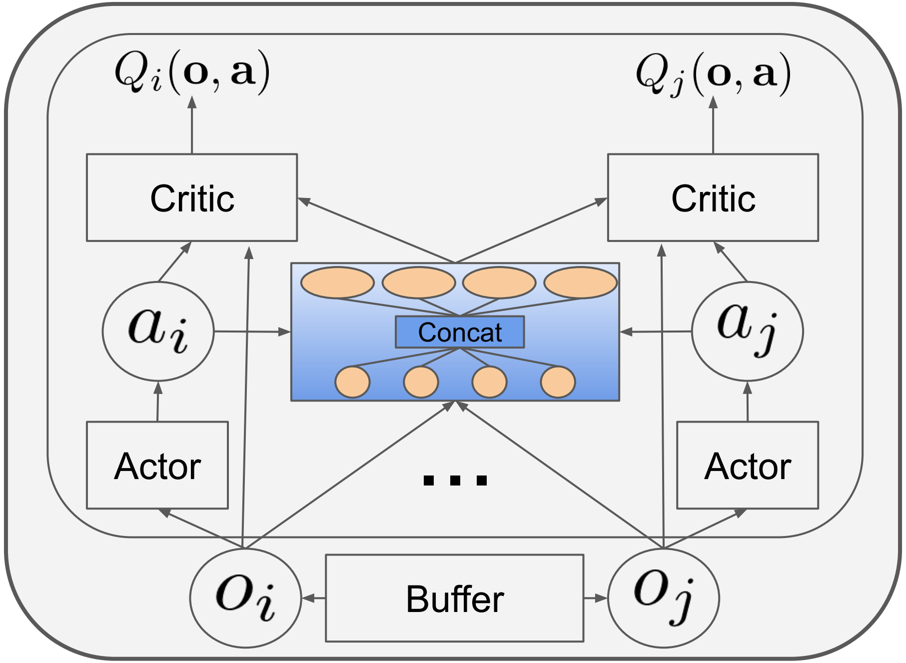
Training

- Centralizing training = each agent's critic takes other agents' actions and observations into account when predicting their own returns
- Policies remain decentralized
- **Pros**:
  - Gives more information to each agent, improving performance
- **Cons**:
  - Now we need communication during training

[1] Foerster, J., Farquhar, G., Afouras, T., Nardelli, N., and Whiteson, S. Counterfactual multi-agent policy gradients. In AAAI Conference on Artificial Intelligence, 2018.

[2] Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, O. P., and Mordatch, I. Multi-agent actor-critic for mixed cooperative-competitive environments. In Advances in Neural Information Processing Systems, pp. 6382–6393, 2017.

But, How to Share?



Training

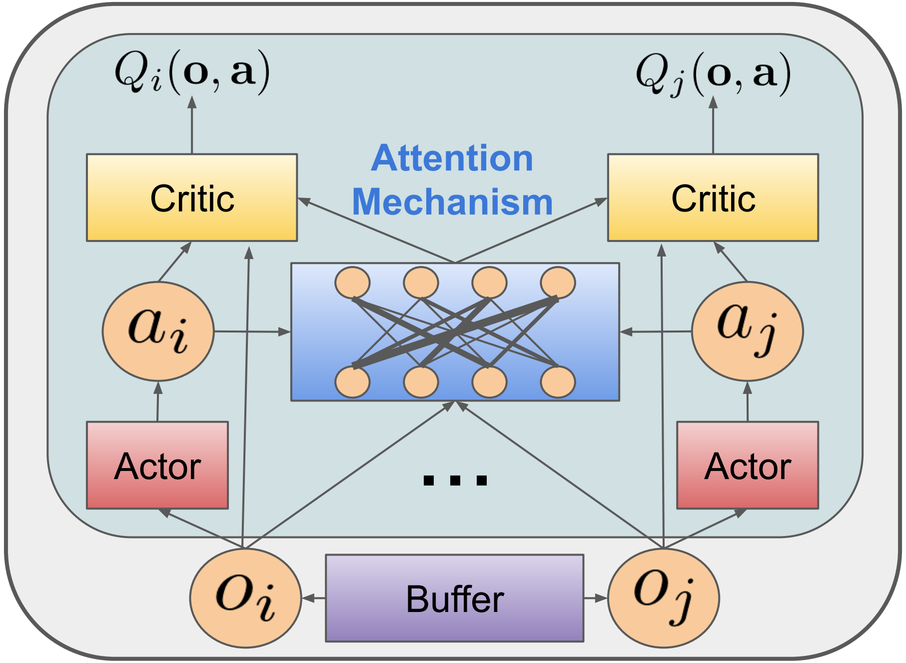
- Existing approaches [1,2] concatenate all information into one long vector
  - Can get large as many agents are added
  - Not all information is relevant

[1] Foerster, J., Farquhar, G., Afouras, T., Nardelli, N., and Whiteson, S. Counterfactual multi-agent policy gradients. In AAAI Conference on Artificial Intelligence, 2018.

[2] Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, O. P., and Mordatch, I. Multi-agent actor-critic for mixed cooperative-competitive environments. In Advances in Neural Information Processing Systems, pp. 6382–6393, 2017.

Actor-Attention-Critic

Sharing information between agents using an attention mechanism

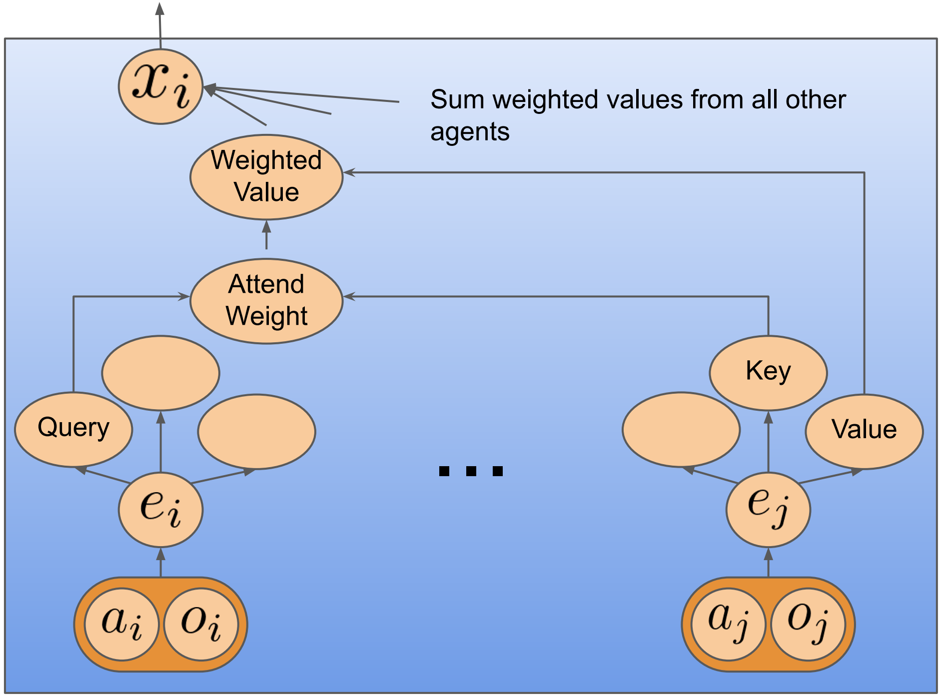


Training

- Agents “attend” to information that is important for predicting their returns
- Information about other agents is encoded into a fixed size vector

Attention Mechanism in Detail

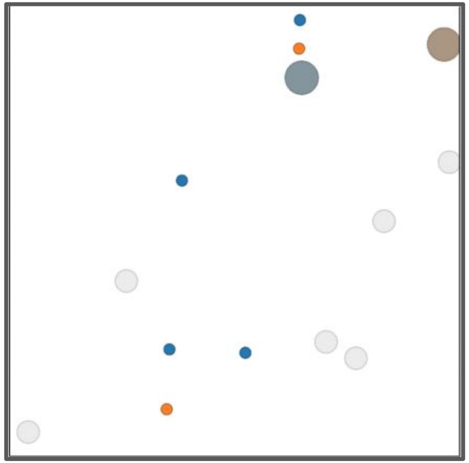
Sharing information between agents using an attention mechanism



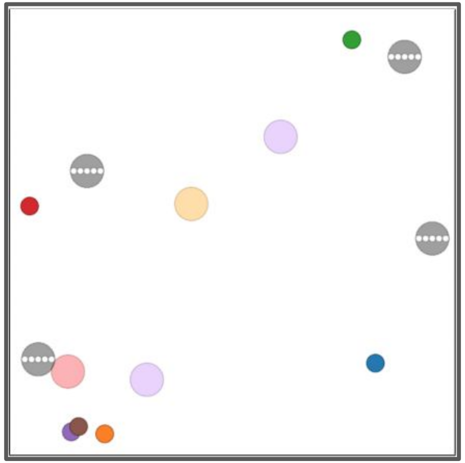
Attention Mechanism

- Agents exchange information using a query-key system
- Ultimately receive aggregated information **from other agents** that is most relevant to predicting their own returns

Environments



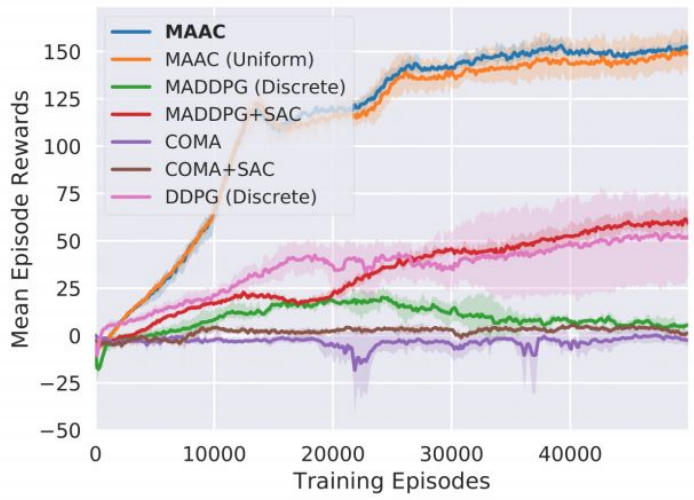
Cooperative Treasure Collection



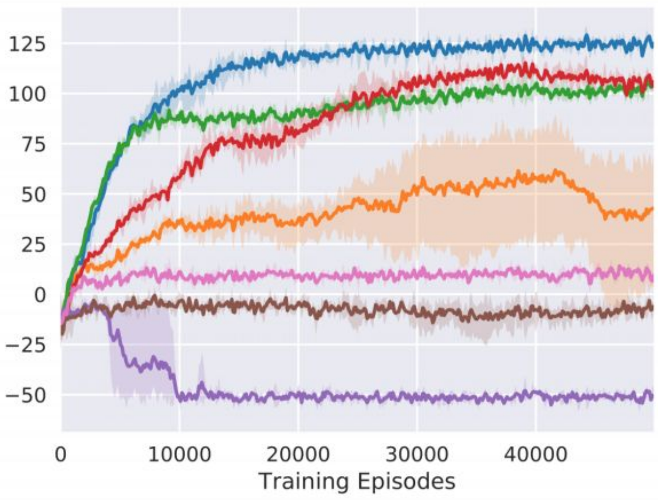
Rover-Tower

- Cooperative Treasure Collection
  - Agents with different roles cooperate to collect colored “treasure” around the map
  - **Challenge**: rewards are shared, and agents must perform multi-agent credit assignment
- Rover-Tower
  - Blind “rovers” and stationary “towers” randomly paired and must cooperatively reach goal through communication
  - **Challenge:** rewards are independent per pair, so agents must learn to select relevant information
- Both tasks are easily scalable and require coordination between heterogeneous agent types

Performance



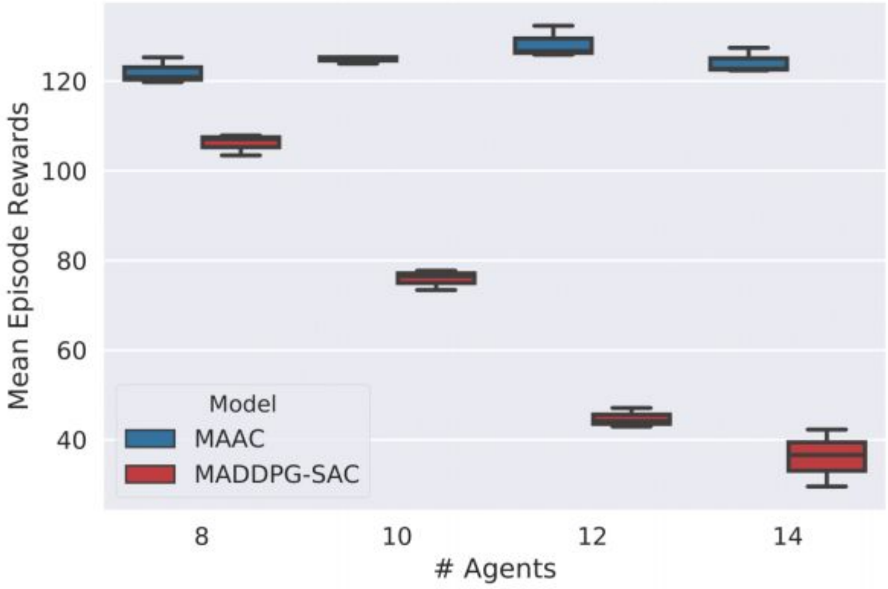
Cooperative Treasure Collection



Rover-Tower

- Our method outperforms baseline methods on two cooperative tasks

Scalability



Rover-Tower

- Compared to the next best performing baseline, our method scales well as agents are added

# Agents	4	8	12
% Improvement	17	98	208

Cooperative Treasure Collection

Thank you!

For more details please come to our poster:

06:30 -- 09:00 PM Pacific Ballroom