Deep Multi-Agent Reinforcement Learning

Liu Yang, Pei-Chen Wu, Guanzhi Wang {liuyeung, pcwu1023, guanzhi}@stanford.edu

Project Summary

Problem:
Explore multi-agent reinforcement learning for a strategy game, StartCraft II, with focusing on micromanagement challenges

Why is this important and why MARL?
- Many real-world problems that might be tackled by RL are inherently multi-agent in nature
- Deep RL promises a scalable approach to solve arbitrary sequential decision-making problems

Challenges:
- Scalability
- Partial Observability
- Inefficient exploration

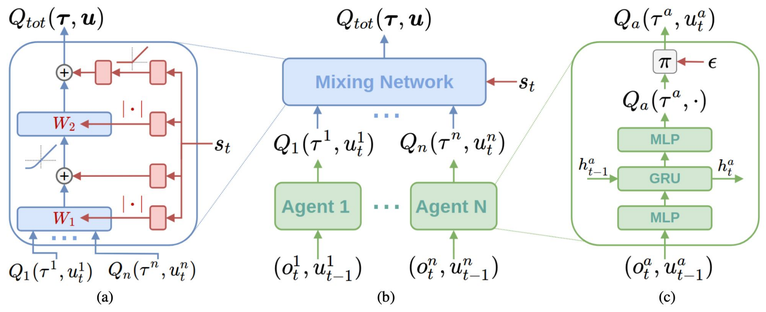
Simulator/Environment:
- SMAC (SC2LE)
- PyMARL

State-of-the-art algorithm:
- QMIX

QMIX Performs badly:
- 5m_vs_6m
- 27m_vs_30m
- 3s_vs_5z
- 2c_vs_64zg

Related Work

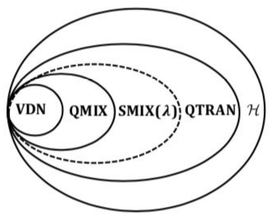
QMIX:
(a) Mixing network structure. In red are the hypernetworks that produce the weights and biases for mixing network layers shown in blue.
(b) The overall QMIX architecture.
(c) Agent network structure.



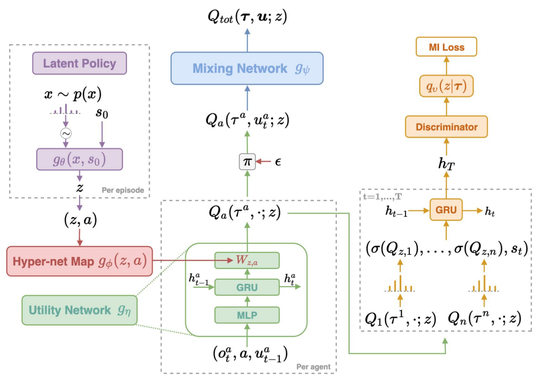
QMIX Architecture

SMIX:
- importance sampling
- λ-return as a proxy
- increases the hypothesis space

MAVEN: learns a diverse ensemble of monotonic approximations with the help of a latent space



SMIX Hypothesis Space

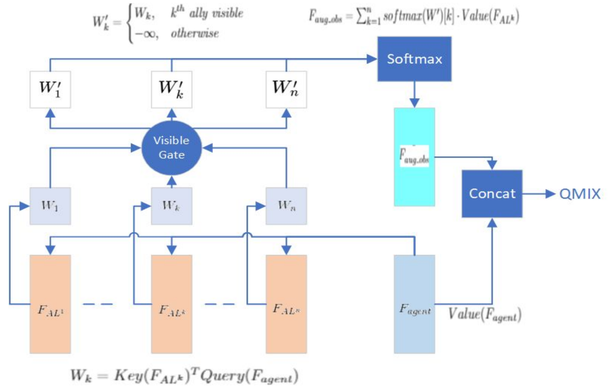


MAVEN Architecture

Methods

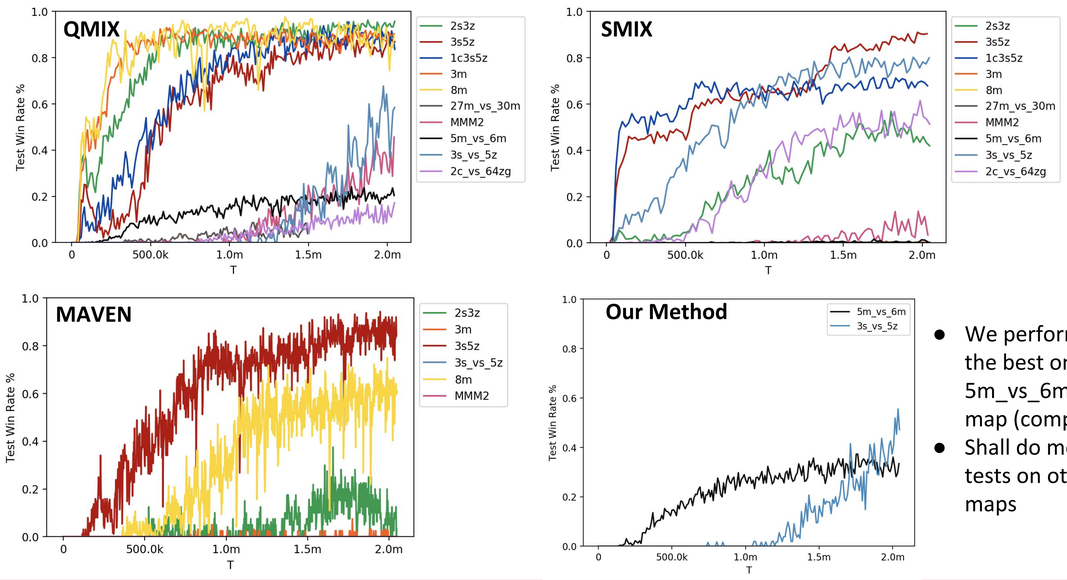
Improvement 1: Communication-Attended Observation

- **Intuition**: Agent's awareness of the environment can be augmented by the observations of its allies within visible range.
- **Attention**: Extract (Query, Key, Value) tuple for observation of each agent. The agent applies Transformer-style attention to allies.
- **Visible Gate:** Only allies within the visible range contribute to the agent's augmented environment observations
- Concatenate the agent's original observation with the augmented one.



Improvement 2: Combine the merits of SMIX and MAVEN (on going). We plan to leverage the expanded hypothesis space of SMIX and embed MAVEN-like latent variables (may be more) to coordinate the actions of agents in latent space.

Experiments



- We perform the best on 5m_vs_6m map (complex)
- Shall do more tests on other maps

References

[1] Rashid, T., Samvelyan, M., De Witt, C. S., Farquhar, G.,Foerster, J., and Whiteson, S. QMIX: Monotonic Value Function Factorisation for Deep Multi-Agent Reinforcement Learning. International Conference on Machine Learning, 2018.
[2] Wen, C., Yao, X., Wang, Y., and Tan, X. Smix(λ): Enhancing Centralized Value Functions for Cooperative Multi-Agent Reinforcement Learning. In Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence, 2020.
[3] Mahajan, A., Rashid, T., Samvelyan, M., and Whiteson,S. Maven: Multi-Agent Variational Exploration. In Advances in Neural Information Processing Systems, 2019.