Structured Multi-Agent World Models

New Jun Jie, Wu Yujin

Problem Statement

In multi-agent reinforcement learning (MARL), the difficulty of generalising to diverse strategies and adapting to non-stationary behaviour remains a challenge. Inspired by model-based reinforcement learning (MBRL), we propose SMAWM or **S**tructured **M**ulti-**A**gent **W**orld **M**odels, a world model that encompasses other agents in a compositional structure, to provide a strong inductive bias to novel interactions among multiple agents in the environment.

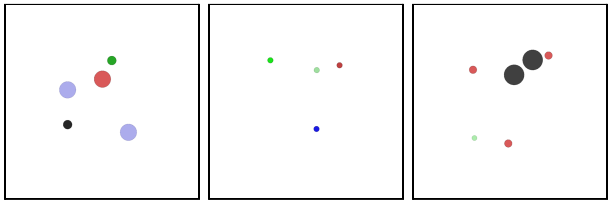
Background

Independent learning that reduces multi-agent learning to single-agent perceives interaction between agents as noise [1]. In contrast, by observing other agents' actions, opponent modelling allows prediction of other agents' future actions, generalising to novel interactions more effectively [2]. Policy reconstruction, one method of modelling agents, would allow for prediction of action probabilities of the modelled agent, improving planning and decision making [3].

Proposed Algorithm

SMAWM is a world model inspired by C-SWM [4], but in contrast, consists of a CNN-based agent extractor, a MLP-based agent encoder and a GNN-based transition model. Our goal is to account for the compositional nature of multiple agents in an environment and learn an agent-factored state space $Z = Z_1,..., Z_k$, where k is the number of agents. The agent-factored state space would serve as a strong inductive bias for better generalization to novel interactions, facilitating planning and decision making by the agent. The transition model predicts the next state z_t given observed agents' actions $A_1,..., A_k$. The agent is trained purely within SMAWM to learn to perform a task, as in World Models (WM) [5].

Task Environment



MPE Tasks: Simple Adversary, Simple Push and Simple Tag

The Multi-agent Particle Environment (MPE) contains a set of tasks where particle agents can move, interact with each other and with fixed landmarks [6]. We choose Simple Adversary, Simple Push and Simple Tag as the tasks are simple and involve cooperation and competition among only a few agents.

Experiment Formulation

Our goal of this research project is to investigate whether SMAWM can model environment dynamics consisting of multiple agents more effectively than naive independent learning, facilitating generalisation to novel interactions and more accurate planning and decision making. We expect SMAWM to outperform WM in sample efficiency and final performance. We also expect SMAWM to outperform WM in the accuracy of prediction multiple steps into the future.

References

[1] Shoham, Yoav, Rob Powers, and Trond Grenager. "If multi-agent learning is the answer, what is the question?." Artificial intelligence 171.7 (2007): 365-377.

[2] Bloembergen, Daan, et al. "Evolutionary dynamics of multi-agent learning: A survey." Journal of Artificial Intelligence Research 53 (2015): 659-697.

[3] Albrecht, Stefano V., and Peter Stone. "Autonomous agents modelling other agents: A comprehensive survey and open problems." Artificial Intelligence 258 (2018): 66-95.

[4] Kipf, Thomas, Elise van der Pol, and Max Welling. "Contrastive learning of structured world models." arXiv preprint arXiv:1911.12247 (2019).

[5] Ha, David, and Jürgen Schmidhuber. "World models." arXiv preprint arXiv:1803.10122 (2018).

[6] Lowe, Ryan, et al. "Multi-agent actor-critic for mixed cooperative-competitive environments." arXiv preprint arXiv:1706.02275 (2017).