Proposal

Notes

This work compares emergent discrete communication—using quantized message protocols—and continuous communication, both managed through Graph Attention Networks (GAT), for complex multi-agent reinforcement learning (MARL) tasks built on the QMIX framework. By integrating GATs to govern which agents communicate and how, the study systematically evaluates how discrete channel constraints (e.g., using vector quantization) versus raw continuous message sharing affect coordination, learning efficiency, and overall task performance in challenging, cooperative environments.

MARLo Polo: Emergent Communication in Multi-Agent Reinforcement Learning for Complex Tasks

1 Introduction

Since Google Deepmind released AlphaGo in 2016, there has been a steady increase of interest into the field of reinforcement learning. This has led to breakthroughs in areas of finance, trading, robotics and even video games. A clear next step for reinforcement learning is multi-agent reinforcement learning where multiple agents are placed into the same environment and have to learn to interact with one another efficiently. This area of research has potential to improve surgery robots, traffic optimisations and search and rescue operations. However, there exist a range of challenges which make multi-agent reinforcement learning more difficult in practice such as: non-stationary problem, credit assignment issues, and the one I will focus on in this project, communication problems. Unlike single-agent reinforcement learning, multi-agent reinformcent learning has the ability to communicate to improve the agent's knowledge on the environment and/or teach them the optimal next step. A very common issue with communication is the "Joint Exploration Problem", this is a struggle for agents to communicate effectivly in novel environments. A recent paper[1] described a technique which is able to counter this issue was suggested.

In this project, I propose to extend AI Mother Tongue to a realistic situation for a more rigourous evaluation of the paper's proposed method. I will compare the emergent communication to a simplistic communication protocol for a quantitive comparison.

Subgoals focusing on larger goals instead of technical goals. Like We must produce a simulation to test our protocols on which can swap communication protocols.

2 Structure of the Project

There are five main components to this project:

- Simulation: Building
- Value Decomposition Network
- Communication Protocols
- Training
- Evaluation: refer to Evaluation section

Simulation

Value Decomposition Network

Communication Protocols

Training

3 Evaluation

Quantitive

Qualitive

4 Starting Point

5 Success Criteria

6 Possible Extensions

6.1 HQ-VAE

6.2 GAT

7 Timetable and Milestones

Weeks 1 to 2

Proposal submitted

Weeks 3 to 4

Weeks 5 to 6

Weeks 7 to 8

Weeks 9 to 11

Weeks 12 to 13

Weeks 13 to 15

Weeks 16 to 17

Weeks 18 to 19

Weeks 20 to 21

Weeks 22 to 23

Weeks 24 to 25

Weeks 26 to 27

Weeks 28 to 29

8 Resource Declaration

I will be using my personal desktop computer (AMD Ryzen 7 3700X, 32GB RAM, NVIDIA RTX 5060 Ti) as my primary machine for software development. As a backup, I will use my personal laptop and resources provided by SRCF (student run computing facility). I will continuously backup my code and dissertation with Git version control onto GitHub.