Karlsruhe Institute of Technology

Swarm Reinforcement Learning with Graph Neural Networks

Bachelor's Thesis
of

Christian Burmeister

KIT Department of Informatics
Institute for Anthropomatics and Robotics (IAR)
Autonomous Learning Robots (ALR)

Referees: Prof. Dr. Techn. Gerhard Neumann
Prof. Dr. Ing. Tamim Asfour

Advisor: Niklas Freymuth

Duration: Juli 17st, 2021 — January 17st, 2022

Erklärung

Ich versichere hiermit, dass ich die Arbeit selbstständig verfasst habe, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, die wörtlich oder inhaltlich übernommenen Stellen als solche kenntlich gemacht habe und die Satzung des Karlsruher Instituts für Technologie zur Sicherung guter wissenschaftlicher Praxis beachtet habe.

Karlsruhe, den 17. January 2022

Christian Burmeister

Zusammenfassung

Einseitige deutsche Zusammenfassung (*Abstract*) der Abschlussarbeit. Unabhängig von der Sprache der Abschlussarbeit *muss* eine deutsche Zusammenfassung verfasst werden.

Abstract

The one-page abstract of the thesis.

Table of Contents

Chapter 1.

Introduction

pages: 0.9-1.2 Stuff to talk about in introduction.

- Applications or real-world problems that require a solution.
- MARL: Good for certain applications like
- GNN more and more popular
- Using GNN for MARL
- What has research focused on?
- Some examples from research what can be done.
- What we set out to do what our basic goal was, that should be a natural conclusion of what we talked about above.
- Last item: What the chapters will talk about, what we will talk about the structure of sections
- Structure:
  - Applications or real-world problems that require a solution.
  - Recent applications and research in MARL and GNN
  - What is GNN, What is MARL, what can GNNs it do for MARL? (the main thing we want to talk about, more conceptually)
  - What is my approach I want to talk about here? What was our goal?
  - My work relative to other work. What has other research focused on?
  - talking about the structure of the thesis

Chapter 2.

Preliminaries

pages: 3.6-4.8 This chapter will introduce the necessary concepts that need to be understood. The baseline is a bachelor's degree in computer science without any assumptions made about the elective studies. Topics:

- RL
  - MDP
- MARL
  - PoMDP
- NN
- vanilla message-passing GNN

Chapter 3.

Related Work

pages: 3.15-4.2 20 referenced papers. 2-3 sections

- RL
  - Swarm RL (max, Robin)
  - PPO
  - TRL
- GNN
  - GNNs
  - GATs
  - MeshGraphNets

Deisenroth et al. (2013)

Chapter 4.

Swarm Reinforcement Learning with Graph Neural Networks

pages: 4.26-5.68 This Chapter is moreso a deep dive into the actual solution of the Swarm RL with GNN Algorithm. Our Architecture and stuff.

4.1. Definition of the Problemdomain

Chapter 5.

Experiments

pages: 2-2.68

5.1. General Setup

Talk about my code base what it is based on etc. What I use.

- Optuna
- DAVIS
- Code from: Bayesian and Attentive Aggregation for Multi-Agent Deep Reinforcement Learning

5.2. Tasks

Structure:

- Goal.
- Basic Environment Structure (Torus,)
- Visual: example task completion, with timesteps and total environment
- Agent-Model (Dynamics), Evader-Model (Strategy)
- Reward
- Observation => Data, Culling and Graph

- Rendezvous
- Dispersion
- Single Evader Pursuit
- Multi Evader Pursuit

5.3. Experiments

Let those experiments run over all environments where applicable.

- PPO Hacks vs No Hacks
  - Environments: Rendezvous
  - value-function-clipping (0.0 - 1.0), 1.0 = no clipping
  - normalize rewards
  - ?reward-clipping: graph-normalized constructor: reward-clip = 5, currently no parameter
  - observation-normalization
  - global gradient clipping: max-grad-norm
  - ?tanh (insted of LeakyReLU)
- Number of Hops:
  - Environments: Rendezvous, Pursuit-Multi
  - Environment: Culling Methods: more culling vs less culling, num-agents and dynamics?
  - Network: num-blocks, latent-dimension?, aggregation-function?
- het-neighbor-aggregation: aggr(aggr()) vs conat(aggr())
  - Environments: Pursuit-Single, Pursuit-Multi
  - Environment: Base-Pursuit-Multi with 3+ Hops?
  - Network: latent-dimension, aggregation-function, num-blocks
- random number of agents
  - Environments: Rendezvous
  - Environment: Rendezvous: Culling Methods: more culling vs less culling
  - Network: latent-dimension, num-blocks
- random number of agents + random number of evaders
  - Environments: Pursuit-Multi
  - Environment: Multi-Pursuit: Culling Methods: more culling vs less culling
  - Network: latent-dimension, num-blocks
- dispersion: reward-type and aggregation-function
  - Environments: Dispersion
  - Environment: Culling Methods: more culling vs less culling
  - Network: latent-dimension, aggregation-function
- ?pursuit: reward-type???
  - Environments: Single-Pursuit
  - Environment: nothing?
  - Network: latent-dimension, aggregation-function

Chapter 6.

Evaluation

pages: 7.14-9.52

6.1. PPO Hacks vs No Hacks

- value-function-clipping (0.0 - 1.0), 1.0 = no clipping
- normalize rewards
- reward-clipping: graph-normalized constructor: reward-clip = 5, currently no parameter
- observation-normalization
- global gradient clipping: max-grad-norm
- tanh (insted of LeakyReLU)

6.2. Number of Hops

- Environment: Culling Methods: more culling vs less culling, num-agents and dynamics?
- Network: num-blocks, latent-dimension?, aggregation-function?

6.3. Neighbor Aggregation Type

aggr(aggr()) vs conat(aggr())

- Environment: Base-Pursuit-Multi with 3+ Hops?
- Network: latent-dimension, aggregation-function, num-blocks

6.4. Randomized Number of Agents and Evaders

random number of agents

- Environment: Rendezvous: Culling Methods: more culling vs less culling
- Network: latent-dimension, num-blocks

random number of agents + random number of evaders

- Environment: Multi-Pursuit: Culling Methods: more culling vs less culling
- Network: latent-dimension, num-blocks

6.5. Dispersion

- Environment: Culling Methods: more culling vs less culling, reward-type
- Network: latent-dimension, aggretation-function

Chapter 7.

Conclusion and Future Work

pages: 0.9-1.2 Some introductory paragraph.

7.1. Conclusion

Your conclusion.

7.2. Future Work

- TRL
- More complex task like Box-Clustering
- Transfer Learning for GNNs?

Bibliography

M. P. Deisenroth, G. Neumann, and J. Peters. *A survey on policy search for robotics*. now publishers, 2013.

Appendix A.

Example Appendix

This is an example for an appendix.