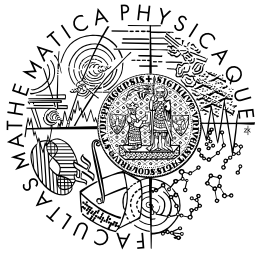


FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Name Surname

Thesis title

Name of the department

Supervisor of the master thesis: Supervisor's Name
Study programme: study programme
Study branch: study branch

Prague YEAR

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In ..... date ....

Author's signature

Dedication.

Title: Thesis title

Author: Name Surname

Department: Name of the department

Supervisor: Supervisor's Name, department

Abstract: Abstract.

Keywords: key words

Contents

Introduction

1. Gentle summarization of Reinforcement learning

An example citation: Anděl [2007]

1.1 Defining mathematical common ground for environment

2. Policy gradient methods

2.1 Idea, motivation and brief technical description of algorithm

2.2 Variants of policy theorem

Vanilla

PPO

3. Multi agent environments for RL

3.1 Definitions

3.2 Possible mention of MAPPO success

4. Overcooked environment

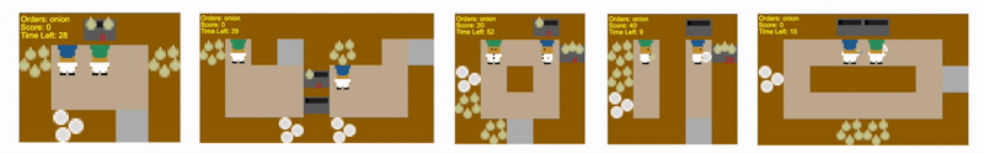
4.1 Overcooked game

Before we get into our problems with cooperation let us first examine the environment. We will be working with environment based on popular cooking video game `https://ghosttowngames.com/overcooked/`. Overcooked is multiplayer cooperative game where the goal is to work in a kitchen as a team with partner cooks and prepare together various dishes within limited time. However, the game is dynamic to a great extent. In many maps the kitchen itself is not static and may be changing on a run. Moreover, random events such as pots catching fire add to the chaos. The challenge lies in coordination with rest of the team and dividing subtasks efficiently.

The aforementioned game was simplified and reimplemented to simpler environment `https://github.com/HumanCompatibleAI/overcooked_ai` to serve a purpose of scientific common ground for studying multi agent cooperation in somehwat complex settings. Lot of additional features of original game were removed and remained only essential aspects. In its simplest form, environment is taking place in small static kitchen layout where only available recipe is onion soup which can be prepared by putting three onions in a pot and waiting for given time period. Somewhere in the kitchen there is unlimited source of onions and dish dispenser, where player can grab a dish to carry cooked onion soup in to the counter. Team of cooks is rewarded as team by abstract reward of value 20 every time cooked soup is delivered to the counter. It may seem that the task is quite straightforward. However, players face problems on multiple levels.

4.2 Basic layouts

Although the Overcooked implementation has its own generator that can be used to generate new random kitchen layouts, the majority of the related scientific work has so far experimented with a fixed set of predefined layouts. Where each layout captures some important aspect of coordination.



(From left to right: Cramped room, Assymetric advantages, Coordination ring, Forced coordination, Counter circuit)

Cramped room as a name suggests represents cramped kitchen layout where all important places are relatively easy to reach. Challenge lies in low level coordination of movement with the other partner as there is no spare room.

In Assymetric advantages both players are located in separated regions where each region is fully self-sustaining. However, each region has better potential for specific subtask. And it is only when both players make the most of their region's potential that the maximal shared efficacy is reached.

TODO: to be continued

4.3 Problem of robustness

Definition of robustness problem

Ad hoc agent playing? Trivial states failure?

4.4 Human cooperation vs artificial cooperation

While it is more intriguing to study consequence of human-ai cooperation. This won't be our main point of focus, since experimenting with humans requires non-trivial overhead of human results evaluation.

5. Lack of robustness problem of AI agents

5.1 Related work

Common approaches fail when paired with foreign

6. Our contribution

6.1 Utilized framework

Comparision of rllib and StableBaselines3

Rllib framework was used in original paper, however for our usages stable baselines seemed sufficient and reasonably easy to extend. Stable baselines has no explicit support of multi-agent environments.

Essential modifications of stable baselines

By default SB3 comes with various wrappers to support most of common environment settings, including

Essential modifications of stable baselines

Initializing methods that resets environment did not yield correct initial positions for some maps.

6.2 Robustness metric

Defining our own metric for robust cooperation of two sets of agents. Probably just average of pair results (non diagonal in case of same sets). Maybe number of pairs who surpassed some threshold?

6.3 Population build up

6.4 Population policies difference rewards augmentation

6.5 Population policies difference loss

Conclusion

Bibliography

J. Anděl. *Základy matematické statistiky.* Druhé opravené vydání. Matfyzpress, Praha, 2007. ISBN 80-7378-001-1.

List of Figures

List of Tables

List of Abbreviations

A. Attachments

A.1 First Attachment