Danmarks
Tekniske Universitet

Causal Graphs in Reinforcement Learning

Project Plan

Author:
Magnus Fredslund s183905
Jakob Andersen s183914

March 12, 2021

1 Topic

Inference of causal representations is an important and difficult problem in Machine Learning and AI. AI is widely used, but is often characterized by being a "black-box" algorithm, where it's hard to explain exactly how the AI works. This problem will focus on identifying causal variables and their relationship in a Reinforcement Learning setup (Causality here is understood as the static definition characterized by for example Judea Pearl [1]).

The goal is to do some kind of temporal abstraction on the states, in order to get information about the underlying causality in the environment, which can be used to improve the AI, and be a stepping-stone to other domains such as high level planning and AI's which uses counter-factual questions.

With a causal variable we mean a collection of states which stand in causal relationship to each other, such a variable could be reward or other similar states which characterizes the environment [2]. We will propose and investigate our own idea, which directly estimates causal states and their relationship from Reinforcement-learning data.

2 Methods

Our main methods is different combinations of reinforcement learning using Q-learning and some other ideas inspired by causality principles. We are using Q-learning to train two different agents which operates on different temporal abstraction levels of the game. The setup is similar to the one in [3], where a meta-agent decides which policy is used in a state.

Our first agent, called the Mover, is playing on a modified representation of the environment. Every state the Mover sees is augmented with a flag somewhere, and the agent only gets reward if it reaches the flag. Our second agent, called the Teleporter, is given the true state, and has to place a flag somewhere in it. The Teleporter gets the true reward from the environment. The Teleporter only sees states where it has to put a new flag on the environment, and thus from it's perspective it teleports the Mover somewhere whenever it takes an action. The Teleporter is interested in completing the level in as few placed flags as possible, and the Mover is interested in capturing the flags as fast as possible. This gives the Teleporter a more temporal abstracted game, which can be used for causal analysis and planning.

Our experiments are done on different playground pixel-based environments, which will be chosen with the intention of making it possible to compare the found causal structure with the actual dynamics in the environment.

3 Environments

For the testing phase, we created our own environments. In order to do this we created an game engine in python that allow us to easily create new objects (Coins, Walls, Keys...). These object are used to create different procedurally generated levels, where the agent trains. Currently we have generated 2 environments where we can train our agent:

Maze: The player has to navigate a maze, to find the goal. On the way the agent can interact with several objects:

- Keys that unlock doors
- Put ice cream in its cone
- Collect coins

Dirt and Rocks: The player has to get to the goal, while digging dirt and moving rocks to find a path to the goal.

These environments are controllable such that not all the objects occur in the environments (Maze without coins). Furthermore, new environments can be created if needed for testing other causal methods.

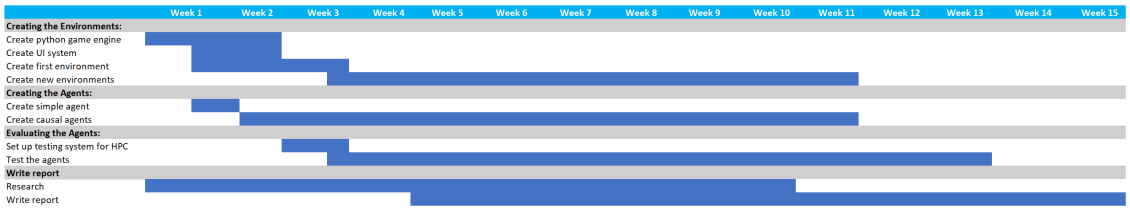
4 Status of knowledge

What do we know about this topic already? Our current knowledge is based on a mixture of the previous mentioned papers, especially [2] and [3]. Furthermore, our testing of some of the environments, and the method described above, have given us a better understanding of the possibilities and the considerations of the project.

5 Outline

1 Introduction
  1.1 State of art
2 Environments
  2.1 Game engine
    2.1.1 UI
    2.1.2 Representation for agent
  2.2 Deception of different environments
    2.2.1 Maze
    2.2.2 Dirt and Rocks
    2.2.3 ...
3 Methods
  3.1 Q-learning
    3.1.1 Mover
  3.2 Causal methods
    3.2.1 Teleporter
    3.2.2 ...
4 Results
  4.1 Results of all methods
5 Discussion and conclusion
  5.1 Difference to non-causal methods

6 Gantt Chart



References

[1] Judea Pearl et al. Causal inference in statistics: An overview. *Statistics surveys*, 3:96–146, 2009.

[2] Rodrigo Toro Icarte, Ethan Waldie, Toryn Klassen, Rick Valenzano, Margarita Castro, and Sheila McIlraith. Learning reward machines for partially observable reinforcement learning. *Advances in Neural Information Processing Systems*, 32:15523–15534, 2019.

[3] Pierre-Luc Bacon, Jean Harb, and Doina Precup. The option-critic architecture. *CoRR*, abs/1609.05140, 2016.