Abstract fiction learning and curiosity approach for general learning

Background

RL good, but not at transferring learning:

Reinforcement learning is already capable of surpassing human capability in different tasks such as Go.

The strength of reinforcement learning is that, it doesn’t require any background information about the task, it just interacts with 3 components of task (observation, action space, reward feedback), and with enough trial and error, it can figure out novel approaches to the challenge that even humans don’t know of. This unsupervised nature makes reinforcement learning exciting, but the very nature of this approach makes skill transfer hard. Since the agent doesn’t have understanding what abstract idea the state and action sets are representing, it fails to use learning of a game to another similar game if the 3 components (observation, action space, reward feedback) of that game are changed. And thus this approach fails to generalize intelligence.

Why are humans good at transfer learning?

Reason 1: diverge representation of same challenge

Humans on the other hand, are usually really good transfer learning. We require very little data to learn to solve a novel challenge compared to reinforcement learning agents. A major challenge of games are that, game representations are often creative, and evolving. Different games might depict same kind of object in diverse ways, which can even be unrealistic. For example – the turtle like creatures in Mario game represents negative rewarding objects. If similar looking turtles are depicted as positive reward in another game, a reinforcement learning agent trained in Mario game will not be able to transfer its learned features in that new game.

Humans are on the other hand very good at fiction association (Harari). A Mario player will have no problem associating negative idea with turtles in Mario game, and positive idea in another situation. In fact, we do this all the time, we’ll associate great value to a piece of paper called money, and get scared of scenes of a horror movie that we know can’t harm us.

Reason 2: state consciousness

Humans also usually understand the challenges that are represented by the system of state objects of a game. Usually, the games directly or indirectly depicts human life – from state representations that includes path, car, house, objects to actions such as movement, driving car, collecting reward etc. This similarity helps us understand the nature of the challenge, and probably solution approach. Because of such consciousness, we require little information to solve the challenge.

But reinforcement learning agents don’t have such state and action consciousness, thus learning good features and appropriate actions in a complex state is computationally expensive.

THE cops 
0.97 
1,349 YARDS 
Il?ll 
10 

In the figures above, two different racing game is depicted, that features different looking state, different action space, different reward structure etc. But the underlying theme is same, challenge of driving a car well. Human players won’t have much problem realizing the similarity of ideas the two games are representing, and in reasoning about possible actions to take – controlling movements of the car, avoiding obstacles etc. Learning to play one game will help the other game to some extent and vice versa. But a reinforcement learning agent have to learn them from scratch, as they can’t understand similarity in the ideas the states are representing.

Why generalization is important?

Generalization of learned skills for artificial intelligent agent is crucial, because real life challenges often features novel complex challenge, and being able to use past learnings in new situation is crucial. Since artificial intelligent agent don’t have limitation of computation, storage or power capacity like we humans do, theoretically an artificial agent has no limits to learn and transfer skills.

Idea

The purpose of this work is to discuss Abstract learner – a method to make learning process of reinforcement learning more general.

Part 1: Coherent representation of reward, death and obstacles

This paper proposes a two part model for reinforcement learning agent. In addition to core reinforcement learner, there will be an additional abstract idea learner process. The abstract idea learner will have three functions, it’ll detect reward nature of the state objects against action system, what results the actions are causing, and it’ll represent the objects with similar reward nature of different tasks in a coherent fashion.

There’ll be a core reinforcement learner agent – ‘Master Learner’, that has a general action space, and that solves task with general intrinsic reward.

The abstract learner will receive observation and action space from different tasks, and it’ll represent those observations in a general way in unsupervised fashion, and learn action sub policies and link them to master actions – it’s like making different tasks different levels of the same ‘Master game’.

The assumption here is that, if objects with similar reward nature across different games are represented to the reinforcement learner agent in a coherent fashion, then the agent will be able to learn features that are more general and transferable to similarly represented new challenges.

Part 2: generalizing intermediaries

Problem with expressing with reward for possible action system

Learning about purpose of different objects in complex state

Using the reward learner for action system, it should be possible to represent positive and negative reward object and obstacles coherently. But in complex games with sparse reward, the intermediate objects have reward of ‘0’. Expressing them in terms of just rewards won’t preserve the context information of these objects.

Problem of keeping B&W: lack of coherence

One way to preserve the information these objects are representing is to keep the black and white pixel values. But in this case, the representation of objects with similar purpose but different appearance won’t be coherent. And the benefit of expressing objects with normalized reward value and possibility of action will be more and more trivial with increasing complexity of games. For example, for normal path of movement, just the information of ability to move might be sufficient, but for more complex objects like the key in the game 'mont rev', more information needs to be preserved in a coherent fashion so that the agent both learns important feature from the state, and is able to transfer the learning in similar but new situation.

Another possibility: master game embedding

If there was a master game that incorporated all possible state challenge unit, then it might have been possible to design an agent that learns to play the game, and embed the learnings in a continuous space. Then in a new game, a similar state object could be expressed with general numeric value from embedding space.

No master game > human transfer learning > representing with location of embedding space

But there is no such master game till now to draw learnings from. It is possible to transfer human learning to the agents. Let's assume we're using word embedding of English language as a part of the memory of the agent. We then use supervised image segmentater to classify intermediary objects of a game that resembles real life objects. For example, if the agent learns that an object in the 'mont rev' game is a key, then the agent could search in its embedded memory for key, and use the normalized location value as representation of the key instead of the raw pixel value.

Another benefit: solution reasoning ability

This would also allow the agent to search for nearest action verbs and properties of the object to reason the solution of the task. For example, let’s consider an embedding where only the master action related verbs are included in the embedding memory. In such case, upon discovering a ladder, the agent could've found that the nearest associated master verb is ‘vertical position change’, and the information would help the agent reason about the task. The agent could similarly learn if the objects embedding was more close to ‘opportunity’ or ‘threat’. The agent could thus figure out diverse information about the state object in a general fashion with just the information of where the object is in the embedding space.

Overall: transferability

And for different games that resembles real life resembling objects, the agent could express different states in a general fashion, and the core agent can learn features that can be reused in another game with similar objects. Furthermore, this can open the door for machine learning agent not just transfer learnings among themselves, but also transfer from and to human understanding.

Problem of using human knowledge

However, the problem of using human knowledge is that the agent can be constrained and biased by human learning.

Model intuition

Reward nature learner

Game State as Pixel Input

Action Space

Semantic disentangled segmentation

Game

Empty matrices for objects

B&W scene

Reward nature learner

Action possibility learner

Embedding location learner

Abstract object representation

New scene representation

General sub policy learner

General action sub policies

Master RL

Model intuition:

State learner:

Semantic segmentation/disentanglement learning

The agent interacts with the game with 3 information – state observation, action space and reward feedback. The agent will take the observation pixels, and use semantic segmentation and disentanglement learning to separate the state objects. The state object will be made black and white, and a new layer with object value placeholder will be created. This can be achieved with semantic segmentation algorithm such as FCN (reference), and disentanglement learning can be used not only for segmenting in unsupervised fashion, but also for separately learning about object units using MoNet/COBRA (reference).

Abstract learner

Three kinds of abstract idea learning process is discussed here. First of all, the learner agent will interact with each object with combinations of action system, and find out if interaction can be carried with that object. In case of obstacles, interaction won’t be possible, but for any other state object, interaction will be possible.

Further, there will be objects that’ll result positive or negative reward, and the normalized values of the object will be recorded.

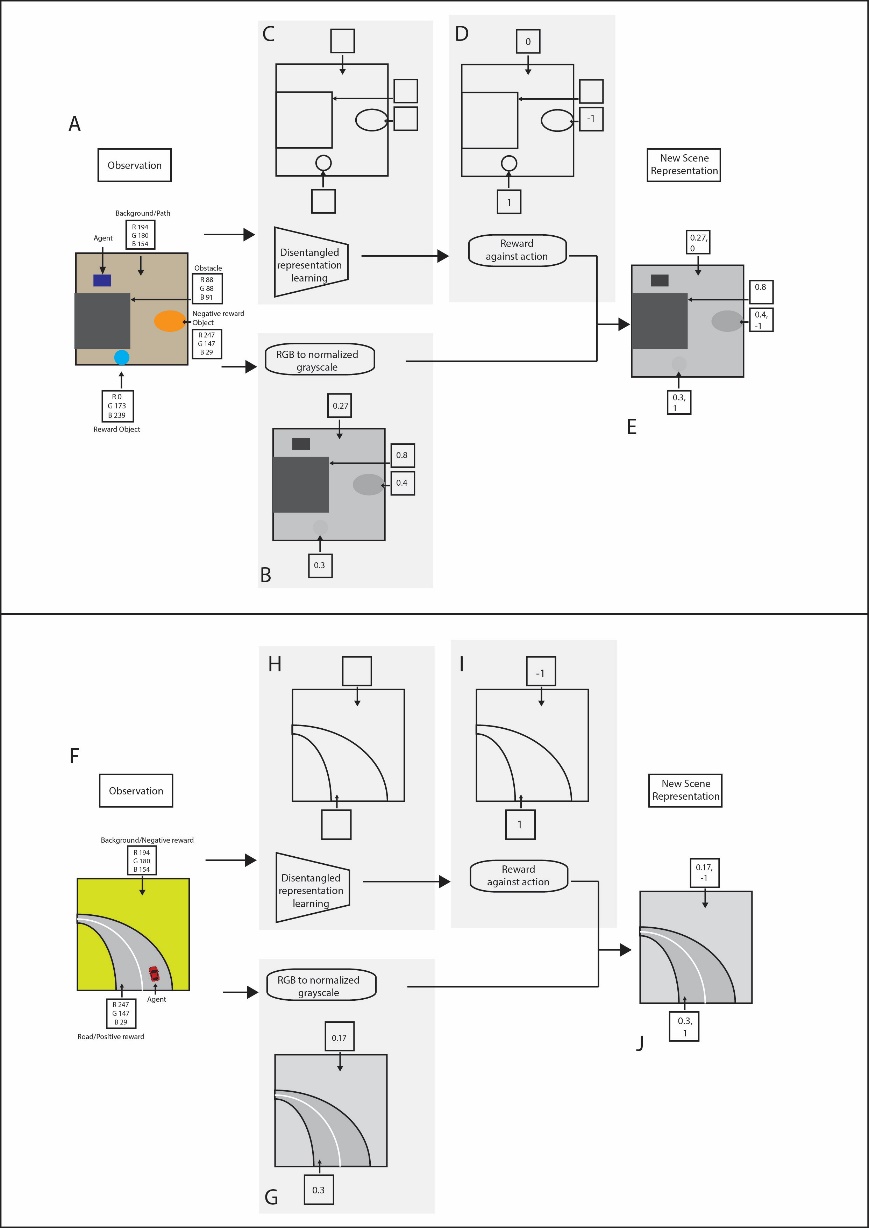
This two learner can be executed with unsupervised disentanglement learner. However, classifying the intermediary objects and expressing them in terms of embedding will require supervised semantic segmentation algorithm stated above.

This entire learner will be executed in exploration focused approach. For example, intrinsic curiosity can be used to learn reward nature by minimizing the predicted reward nature of the object.

The result will be high dimensional but generally represented state. It’ll be appended with the normalized black and white pixel values to create new state representation. This representation will be passed onto a master reinforcement learner, which can learn general and transferable feature form the newly represented state. The master agent will search for nearest master action verbs in the word embedding from memory, and use that information to start learning.

Action learner:

The action learner will take actions from the game action space, and figure out proper set of actions that results in general change in state – horizontal, vertical change in position of agent etc. It’ll learn necessary sub policies that achieves general action result, and the master agent will use those sub policies.



Similar works

IMPALA: IMPALA is perhaps one of the best general reinforcement learning agent. It is an efficient model with separate actor and learner, where the actor takes action and send the learner observation, and the learner updates a common policy. Thus, in a challenge set like DMLab-30, that has 30 different task with common action space, the model learns transfer between the learnings within DMLab-30 tasks.

COBRA: Curious Object-Based Search Agent – COBRA uses data efficient MoNet for unsupervised object segmentation and uses just curiosity based exploration to play games.

SNAIL: Simple Neural Attentive Meta Learner combines temporal convolutions and self-attention to distill useful information from the experience it gathers. It performs well in few shot image classification and reinforcement learning tasks.

Reference