Paper Note

Remosy

A report submitted for the course
COMP8755 AND Individual Computing Project
Supervised by: Dr. Penny Kyburz
The Australian National University

July 2019

© Remosy 2019

DRAFT – 31 July 2019

Except where otherwise indicated, this report is my own original work.

Remosy
31 July 2019

Contents

1 Atari Gym:Ice Hockey 1

1.1 Gym . . . 1

1.2 Ice Hockey . . . 1

1.3 Record:Ice Hockey . . . 1

2 Learning from Demonstration 3

2.1 Playing hard exploration games by watching YouTube . . . 3

2.2 Player Experience Extraction from Gameplay Video . . . 3

2.3 Learning Montezuma's Revenge from a Single Demonstration . . . 3

2.4 Deep Q-learning from Demonstration . . . 3

3 Object Detection 5

3.1 My Application . . . 5

3.2 Object Detection Pre-Train techniques . . . 5

3.3 Object Detection Trainning techniques . . . 5

3.4 Object-oriented state Abstraction in RL for Video Games . . . 5

3.5 YOLO . . . 6

3.6 CenterNet: Keypoint Triplets for Object Detection & Objects as Points . 6

3.7 Object Tracking Vs. Object Detection . . . 6

4 State Embeddings 7

4.1 Dynamic image encoder . . . 7

4.2 Stack of difference of frame video-clip encoder . . . 7

4.3 Grad-CAM ++ . . . 7

4.4 VCG . . . 8

5 RL 9

5.1 Definitions . . . 9

5.2 Off-Policy . . . 10

5.3 On-Policy . . . 10

5.4 Model Based . . . 10

5.5 Model Free . . . 10

5.6 DQN . . . 10

5.7 A3C . . . 10

6 IRL 11

6.1 GAN . . . 11

6.2 Autoencoder . . . 11

6.3 Hidden Markov Model . . . 11

6.4 Apprenticeship Learing . . . 11

6.5 Bayesian Inverse Reinforcement Learning . . . 11

6.6 Maximum Entropy Reinforcement Learning . . . 11

6.7 Generative Adversarial Imitation Learning . . . 11

6.8 Project Scope . . . 11

6.9 Report Outline . . . 11

7 Policy Optimisation 13

7.1 Dynamic image encoder . . . 13

7.2 Stack of difference of frame video-clip encoder . . . 13

Bibliography 15

Atari Gym:Ice Hockey

1.1 Gym

Open AI published a reinforcement learning toolkit, gym. It includes the Arcade learning environment(Bellemare et al., 2012) ran on Stella Atari emulator. Therefore, user can train AI agents for a large number of Atari games from the gym library.

1.2 Ice Hockey

Ice Hockey, an Atari video game released in 1981. It has two game modes: single and 2-player multiplayer. Ice Hockey has two teams, one team wear yellow cloth and green pants, another team wear blue cloth and red pants. Each team has two roles, goalie and offense. In this game, the action control is taken by either goalie or offense of a team, who is closer to the hockey puck. The whole time for a game episode is 3 minutes.

In the gym library, Ice Hockey has an observation shape (210,160,3); it has 3 layers of RGB, each layer is shaped in height 210 and width 160.

Space: Discrete
timesteplimit = maxEpisodeSteps
Trials = 100
Ram version Vs. Non-ram Version
Deterministic? No
NoFrameSkip?

Size = [105 80]

1.3 Record:Ice Hockey

Method 1
Use record wrappers. The

[0] Noop	**[1] Fire**	**[2] Up**	**[3] Right**
-	(32,) ESC	(119,) W	(100,) D
[4] Left	**[5] Down**	**[6] UpRight**	**[7] UpLeft**
(97,) A	(115,) S	(100,119) D,W	(97,119) A,W
[8] DownRight	**[9] DownLeft**	**[10] Upfire**	**[11] RightFire**
(100,115) D,S	(97,115) A,S	(32,119) ESC,W	(32,100) ESC,D
[12] LeftFire	**[13] DownFire**	**[14] UpRightFire**	**[15]UpLeftFire**
(32,97) ESC,A	(32,115) ESC,S	(32,100,119) ESC,D,W	(32,97,119) ESC,A,W
[16]DownRightFire	**[17]DownLeftFire**		
(32,100,115) ESC,D,S	(32,97,115) ESC,A,S		

Table 1.1: Actions

Chapter 2

Learning from Demonstration

2.1 Playing hard exploration games by watching YouTube

2.2 Player Experience Extraction from Gameplay Video

2.3 Learning Montezuma's Revenge from a Single Demonstration

2.4 Deep Q-learning from Demonstration

How many chapters you have? You may have Chapter **??**, Chapter **??**, Chapter **??**, Chapter **??**, and Chapter **??**.

Chapter 3

Object Detection

3.1 My Application

(1)Download youtube Videos via "pytube"
(2)Use "ffmpeg" to make 10-FPS videos
(a)This step can reduce frames of whole videos
(3)Extract video frames from the 10-FPS videos
(4)Annotate video frames by image labelling tools "RectLabel"
(5)Convert VOC format into YoLo format via "RectLabel"
(6)

3.2 Object Detection Pre-Train techniques

3.3 Object Detection Trainning techniques

Q1:Ad hoc training
Q2:How to Classify regions with SVMs
Q3:The use of FC layer
Q4:The use of Spatial Pyramid Pooling (SPP) layer
Q5: What is Linear + softmax
Q6:Region of Interest (RoI) pooling must be (sub′) differentiable to train conv layers
Q7:What can be efficient SGD steps
Q8:What is region′wise sampling
Q9:What is mini batches
Q10:Why is hierarchical sampling help build mini batches
Q11:How Fast-RCNN solve out of region problem?

3.4 Object-oriented state Abstraction in RL for Video Games

Reduce state space size
(1)known coordinates of objects in K classes and demonstrate $\phi : S \mapsto \{0,1\}^{\{L\times L\times X\}}$
(2)divide image s into grid $L \times L$, then represent the K classes by K matrices of size

$L \times L$
Automatic Object Detection
(1) Randomly propose bounding boxes
(2)Autoencoder
(3)k-means
(4)new abstraction ϕ
(5)learn abstracted Q-function
(6)collect new experience states and
(7)update centroids and filter out irrelevant objects
Filter out irrelevant objets by updating centroids
Use k-mean and threshold
Threshold is determined by a rule of thumb based on t-test for Q values **Summary**
Q1:Why should the K in k-means be greater than object classes?
Q2:Why is it a lower dimensional space
Q3:What is the advantages of using k-means
Q4:What is the disadvantages of using k-means
Q5:Why do we use the Q-function on object detection
Q6:What is the "a rule of thumb based on t-test"
Results of this paper

3.5 YOLO

Summary
Q1: What is "num"? What is "anchor"? Why is filter calculated by: num/3 * (class + 5)?
A1:In yolov3, num = 9 Q2: What is weight of a net? What is it for?

3.6 CenterNet: Keypoint Triplets for Object Detection & Objects as Points

Summary
Q1:What drawbacks of CornerNet the CenterNet improved? by what?
Q2:

3.7 Object Tracking Vs. Object Detection

Chapter 4

State Embeddings

4.1 Dynamic image encoder

4.2 Stack of difference of frame video-clip encoder

4.3 Grad-CAM ++

The Class Activation Mapping has been generalised into Gradient Class Activation Mapping (Grad-CAM). The interpretability of Grad-CAM in deep network can lead researchers to failures of model. For the interpretability, the Grad-CAM can discriminate the image classes by localising the region of interest (ROI) in a image.

Grad-CAM uses CNN architecture, the gradient mappings obtained from CNN can be used in the final fully connected layer to significant the regions of interest in a image.

Where to get the gradient information

$$\alpha_k^c = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial A_i^k j}$$

α_k^c :

$\frac{1}{Z}$: global average pooling

y^c :

$A_i^k j$: A matrix of K feature map

$\frac{\partial y^c}{\partial A_i^k j}$:

$$L_{Grad\_CAM}^c = ReLU(\sum_k \alpha_k^c A^k)$$

$ReLU$: it's good to see the positive result

Grad-CAM ++ is better for video-based training.

Why "++" better for video? How differ from Grad_CAM?

4.4 VCG

How many chapters you have? You may have Chapter **??**, Chapter **??**, Chapter **??**, Chapter **??**, and Chapter **??**.

Chapter 5

RL

5.1 Definitions

Agent: An agent takes actions

Action (A): A is the set of all possible moves the agent can make

Environment: The world through which the agent moves. The environment takes the agent′ current state and action as input, and returns as output the agent′s reward and next state

State (S): A state is a concrete and immediate situation in which the agent finds itself

Reward (r): A reward is the feedback by which we measure the success or failure of an agent′ s actions

Discount factor (γ): The discount factor is multiplied with future rewards as discovered by the agent in order to dampen their effect on the agent′ s choice of action. It makes future rewards worth less than immediate rewards

Policy (π): The policy is the strategy the agent employs to determine the next action based on the current state. It maps states to actions

Value (V): The expected long-term return with discount, as opposed to the short-term reward r. $V\pi(s)$ is defined as the expected long-term return of the current state under policy π

Q-value or action-value (Q): Q-value is similar to Value, except that it takes an extra parameter, the current action a. $Q\pi(s, a)$ refers to the long -term return of the current state′ s, taking action a under policy π. Q maps state-action pairs to rewards

5.2 Off-Policy

5.3 On-Policy

5.4 Model Based

5.5 Model Free

5.6 DQN

Use the immediate reward we receive and a value estimate of our new state to update the value estimate of original state-action pair. We only had the learned value function Q-function and the policy we followed was simply taking the action that maximised the Q-value at each step

5.7 A3C

Actor-critic methods combine policy gradient methods with a learned value function. we learn two different functions: the policy (or "actor"), and the value (the "criti"). The "policy" adjusts action probabilities based on the current estimated advantage of taking that action, and the value function updates that advantage based on the experience and rewards collected by following the policy

How many chapters you have? You may have Chapter **??**, Chapter **??**, Chapter **??**, Chapter **??**, and Chapter **??**.

Chapter 6

IRL

6.1 GAN

6.2 Autoencoder

6.3 Hidden Markov Model

6.4 Apprenticeship Learing

6.5 Bayesian Inverse Reinforcement Learning

6.6 Maximum Entropy Reinforcement Learning

6.7 Generative Adversarial Imitation Learning

6.8 Project Scope

Describe the problem your project addresses.

6.9 Report Outline

How many chapters you have? You may have Chapter **??**, Chapter **??**, Chapter **??**, Chapter **??**, and Chapter **??**.

Chapter 7

Policy Optimisation

7.1 Dynamic image encoder

7.2 Stack of difference of frame video-clip encoder

Using a stack of differences of frames to capture motion and dynamics from a video clip. It was helpful to learn odd-one-out

How many chapters you have? You may have Chapter **??**, Chapter **??**, Chapter **??**, Chapter **??**, and Chapter **??**.

Bibliography