Undergraduate Thesis 2021

Title

Supervisor Nakajima Tatsuo
Area of Study Computer Science

Waseda University
School of Fundamental Science and Engineering
Department of Computer Science

1W17BG08-2 Hu Yong-Hao

submitted on 2022.01.31

Abstract

This is my abstract...

Acknowledgements

This is my acknowledgements...

Contents

List of Figures

List of Tables

Chapter 0

Notations

Sample notations ??

Table 1: Mathematical notations

Symbol	Meaning
α	learning rate
γ	discount factor
S, s	state
A, a	action
R, r	reward
τ	a trajectory / an episode
G	return
t	a discrete time step
G_t	return at time step t
T	final time step of an episode
π	policy
π_θ	parametrized policy with parameter θ
$\pi(s)$	the action distribution given state s under policy π
$\pi(a\|s)$	probability of action a given state s under policy π
$\mathbb{E}$	expectation
$\mathbb{E}_\pi$	expectation under policy π
$v(s)$	state value of state S
$v_\pi(s)$	state value of state S under policy π
$q(s, a)$	action value of action a on state s
$q_pi(s, a)$	action value of action a on state s under policy π
σ	activation function

Chapter 1

Introduction

1.1 Motivations

1.2 Objectives and importance

1.3 Overview of this paper

1.4 Sample section

Sample template **alphago**

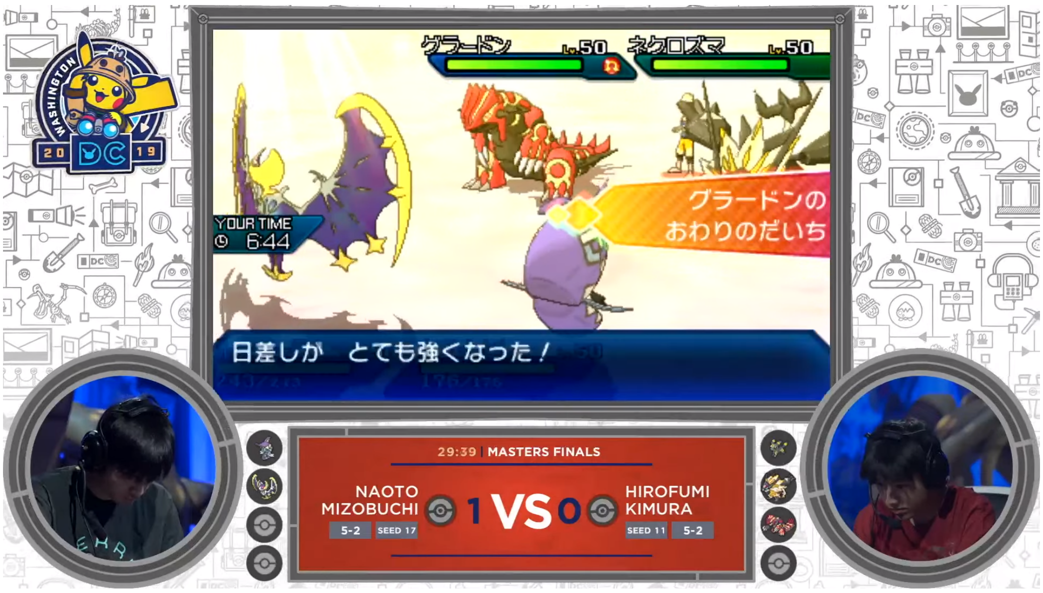

Figure 1.1: Screenshot of the Grand Finals of the Pokemon Video Game Championships 2019 held in Washinton D.C.

Chapter 2

Backgrounds

2.1 Pandemic's impact

2.2 Local transformation

2.3 Location-based Augmented Reality

2.4 Co-creation

Chapter 3

Related Works

3.1 Location-based AR's effect on users' image of a place

3.2 Location-based AR's effect on users' motivation of a place

3.3 Co-creation's effect on users' motivation to access a service

3.4 User-user interaction's effect on users' engagement of a service

3.5 Local transformation by graffiti

3.6 More examples of location-based service with co-creation

Chapter 4

Methodology

Chapter 5

Experiments and Results

5.1 Overview

Chapter 6

Discussions

APPENDIX A - Machine Specs

Table 6.1: Machine specs

Item	Value
CPU	Intel Xeon E5-2690
Memory	188G
OS	18.04.5 LTS (GNU/Linux 4.15.0-121-generic x86_64)

APPENDIX B - Derivation of the simplest form of policy gradient

Derivation of the simplest form of policy gradient is provided below.

$$
\begin{aligned}
\nabla_\theta J(\pi_\theta) &= \nabla_\theta \, \mathbb{E}_{\tau\sim\pi}[R(\tau)] \\
&= \nabla_\theta \int_\tau P(\tau|\theta)R(\tau) \\
&= \int_\tau \nabla_\theta P(\tau|\theta)R(\tau) \\
&= \int_\tau P(\tau|\theta)\nabla_\theta log P(\tau|\theta)R(\tau) \\
&= \mathbb{E}_{\tau\sim\pi}[\nabla_\theta log P(\tau|\theta)R(\tau)] \\
&= \mathbb{E}_{\tau\sim\pi}[\nabla_\theta log \pi_\theta(a_t|s_t)R(\tau)]
\end{aligned}
$$

This is a expectation, which can be estimated with a sample mean. Denote the estimated policy gradient as $\hat{g}$:

$$
\hat{g} = \frac{1}{D}\sum_{\tau\in D}\sum_{t=0}^{T} \nabla_\theta log \pi_\theta(a_t|s_t)R(\tau)
$$