**Nama : Rayinda Maya Anjani**

**Nim : 19SA203**

**Kelas : Ragnarok**

**PERTEMUAN 1**

Modul Domain AI Reinforcement Learning Bagian Introduction to Reinforcement Learning

Pengenalan Reinforcement Learning (RL)

1. Membahas Definisi dan sejarah RL

Karakteristik RL :

• Tidak ada “kebenaran” pada proses pembelajaran (training) yang ada hanya hadiah (atau hukuman)

• Data yang diperoleh (State dan Reward) tergantung dari Action yang diberikan oleh Agent.

• Variabel waktu dan urut-urutan state yang dipilih oleh Agent melalui action memiliki pengaruh

yang penting.

2. Mempelajari Algoritma RL

Objective -> State -> Action -> Reward -> Termination

3. Membahas Element & Environment dari RL

❑ Agent, adalah perangkat atau software (perangkat lunak) yang dapat belajar dari environment.

❑ Environment adalah segala sesuatu yang ada di luar agent yang dimana sebagai tempat agent

untuk melakukan exploration dan exploitation.

4. Membahas Aplikasi dari RL (AI of Game, Robotics, & Optimization)

Reinforcement Learning for AI in Games : Atari Games dan OpenAI Five (Dota)

Reinforcement Learning for Robotics

• Examples of building trajectory generation for Humanoid Robot using Reinforcement Learning.

• Examples of Self Driving Car using Reinforcement Learning.

Reinforcement Learning for Optimization

Contoh dari penerapan Reinforcement Learning untuk Optimasi:

• Menentukan luasan suatu area dengan pendekatan optimasi monte carlo prediction.

• Menentukan model suatu fungsi matematika dengan pendekatan optimasi monte carlo prediction

• Membuat solusi dari permasalahan kasus Traveling Salesman Problem.

• Membuat suatu sistem Simultaneous Locallization and Mapping (SLAM) untuk mengetahui posisi

object terhadap lingkungan dalam koordinat lokal maupun koordinat global.

**PERTEMUAN 2**

Membahas Module RL (Reinforcement Learning) tentang Markov Decision Process (MDP) dan Dynamic Programming

1. Pengenalan Konsep Markov Decision Proses

• Markov Decision Process merupakan sebuah tuple (S, A, P, R, γ). Dimana:

S merupakan state; A merupakan action ; P merupakan state transition probability function (transition

probability); R merupakan reward function ; γ merupakan discount factor (γ ϵ [0, 1])

• Aplikasi RL dalam Dunia Nyata : a) Robotik: Memutuskan bergerak kemana. b) Alokasi Sumber Daya:

Memutuskan apa yang diproduksi c) Pertanian : Menentukan apa yang ditanam, tapi tidak tahu cuaca

dan hasil tanam

2. Belajar menyusun MDP dalam merumuskan RL secara formal

Contoh Transportasi

• Sebuah kota yang mempunyai nomor blok 1 sampai n

• Berjalan dari s ke s+1 membutuhkan waktu 1 menit

• Naik tram ajaib dari s ke 2s membutuhkan waktu 2 menit

• Bagaimana berpindah dari 1 ke n dalam waktu paling sedikit?

• Tram punya kemungkinan gagal 0.5

• Jika gagal, waktu habis 2 menit tapi tidak pindah blok C

Solusi untuk MDP?

• Policy adalah fungsi 𝜋 yang memetakan untuk setiap state 𝑠 ∈ 𝑆𝑡𝑎𝑡𝑒𝑠 ke action 𝑎 ∈ 𝐴𝑐𝑡𝑖𝑜𝑛(𝑠)

3. Menentukan Value sebuah Policy. Berikut Evaluasi pada sebuah policy :

• Mengikuti policy akan bisa menghasilkan jalur yang acak (mengapa?)

• Return (utility) dari sebuah policy adalah Jumlah dari reward selama mengikuti jalur (nilai yang acak)

4. Mencoba membuat Dynamic Programming tipe Value Iteration

Menghitung value secara iterasi

• Algoritme menggunakan iterasi:

• Inisialisasi 𝑉𝜋 0 𝑠 ⟵ 0 untuk semua state

• Untuk interasi 𝑡 = 1,… ,𝑒𝑛𝑑:

• Untuk setiap state 𝑠: 𝑉𝜋 𝑡 (𝑠) ← ෍ 𝑠′ 𝑃 𝑠, 𝜋 , 𝑠 ′ [𝑅 𝑠, 𝜋 𝑠 , 𝑠 ′ + 𝛾𝑉𝜋 𝑡−1 (𝑠′) M

**PERTEMUAN 3**

Membahas Modul Domain AI tentang Monte Carlo Prediction

1. Membahas Monte Carlo Prediction (Konsep MC, elemen MC, Algoritme MC, contoh)

• MC tidak mengambil pengetahuan lengkap dari environment.

• MC belajar dari experience, episode per episode, baik itu experience aktual, maupun simulasi.

• MC belajar dari episode-episode secara utuh dan independent, tidak bootstrapping.

Elemen Algoritme MC

• Goal :: Belajar 𝑣𝜋 melalui episode-episode dari experience yang patuh pada policy

• Return :: Adalah total discounted reward

• Value Function :: Adalah expected return

Contoh

• Misalkan kita punya sebuah environment dengan dua state, yaitu A dan B. Kita telah melakukan

observasi selama 2 episode dan menghasilkan urutan seperti ini :

• A+3 →A+2 artinya adalah transition dari state A →A dengan reward = 3 untuk transisi tersebut.

• Kita akan coba mencari V(A) dan V(B) menggunakan Algoritma MC first visit dan juga every visit.

2. Membahas Monte Carlo Estimation & Control (MS Estimation & MC Control)

MS Estimation : • Karena model tidak tersedia, perlu untuk mengestimasi action juga, selain hanya

mengestimasi state-nya. • Tidak seperti DP, pada Model Free Algorithm, MC, state saja tidak cukup

untuk menentukan policy. • Pada MC, action diperlukan dalam menentukan policy.

3. Membahas On-Policy Monte Carlo (Konsep & algortime)

Konsep : On–policy MC adalah algoritme Monte Carlo yang melakukan evaluasi atau improvisasi dari

policy yang digunakan untuk membuat keputusan-keputusan.

4. Membahas Off-Policy Monte Carlo (Konsep & algoritme)

Konsep : Off–policy MC adalah algoritme Monte Carlo yang melakukan evaluasi atau improvisasi dari

policy yang berbeda dalam meng-generate data (behavior policy), sedangkan target policy-nya sama.

5. Mempelajari Pemrograman Monte Carlo

Kesimpulan :

❑ Monte Carlo bekerja dengan pengalaman dari sample, dan menggunakannya untuk belajar secara langsung tanpa model.

❑ Pada first–visit MC, value function dihitung saat pertama kali mengunjungi state s pertama dalam setiap episode.

❑ Pada every–visit MC, value function dihitung setiap mengunjungi state s dalam setiap episode.

❑ Pada on-policy MC, agent berkomitmen untuk selalu eksplorasi dan mencoba mencari policy terbaik.

❑ Pada off-policy MC, dilakukan evaluasi atau improvisasi dari policy yang berbeda dalam meng-generate data (behavior policy), sedangkan target policy-nya sama.

**PERTEMUAN 4**

Secara garis besar materi yang saya pelajari membahas materi tentang Reinforcement Learning tentang Temporal-Difference Learning adalah sebagai berikut :

1. Recap DP dan MC

2. Membahas Konsep TDL

TDL = Mengestimasi reward pada setiap Langkah (step)

3. Membahas Dynamic Programming dan Monte Carlo Method

Dynamic Programming & Monte Carlo Method Dynamic Programming

❖Update per step, menggunakan bootstrapping ❖Membutuhkan model environment

❖Computation cost Monte Carlo Method ❖Update per episode ❖Model-free (tidak membutuhkan

model environment) ❖Sulit untuk di aplikasikan pd continuing tas

4. Membahas TD Control ( SARSA dan contoh kasus )

SARSA (State-Action-Reward-State-Action) : • Terinspirasi dari policy iteration • Mengganti value

function (𝑉𝜋) dengan action-value function • On Policy • Fakus pada state-action (S, A)

CATATAN : a). Idea dasar dari TDL adalah kombinasi DP dan MC b). TDL melakukan update value state

per step c). Bersifat bootstapping d). Menggunakan metode trial and error dalam explorasi lingkungan

e). TD control menggunakan SARSA dan Q-Learning g). SARSA bekerja berdasarkan perubahan State

Action (S,A) h). SARSA adalah algoritma on policy i).Pengambilan keputusan dari state-action SARSA

berdasarkan epsilon-greedy policy.

• Hands On TD​ dan SARSA

**PERTEMUAN 5**

Membahas Modul Domain AI Reinforcement Learning Belajar dengan Penguatan Bagian Q Learning dan Deep Q Learning

1. Q-learning termasuk dalam kategori Model-free RL algorithm. Menggunakan Bellman Equation untuk perhitungan dan bersifat online action-value function learning dengan exploration policy. Sehingga kita perlu belajar konsep dasar RL sebelum mempelajari Q-learning.

2. Pada Q-learning, agent belajar menggunakan evaluation function yang bergantung pada sekumpulan state dan sekumpulan action.

3. Pada Q-learning dilakukan iterasi: ➢ Policy iteration ➢ Value iteration

4. Di tahap awal, elemen matrix Q-table akan ditentukan bernilai 0 semuanya. Q-table akan di update sejalan dengan bertambahnya jumlah episode.

5. Secara prinsip semakin banyak iterasi episode akan semakin baik Q-table-nya.

6. Q-table ini dapat dianalogikan seperti 'memori' bagi agent. Agent dapat menentukan path optimal berdasarkan nilai yang tertera pada Q-table.

**PERTEMUAN 6**

Secara garis besar materi yang saya pelajari membahas Modul Domain AI Reinforcement Learning tentang Robotics, meliputi sebagai berikut :

1). Mengulas kembali materi

Dynamic Programming (Recap) : Update value state 𝑉 𝑆𝑡 per step. Jadi pada DP algoritma tidak harus menyelesaikan 1 episode secara utuh untuk meng-update value state V(s). Dan bersifat bootstrapping.

Simple Monte Carlo (Recap) : Update value state 𝑉 𝑆𝑡 setelah episode selesai. Jadi pada MC algoritma harus menyelesaikan dahulu 1 episode secara utuh untuk mengupdate value state V(s).

Simplest TD Method (Recap) : Update value state 𝑉 𝑆𝑡 per step seperti dynamic programming. Jadi pada TD algoritma tidak harus menyelesaikan 1 episode secara utuh untuk meng-update value state V(s).

Temporal Difference Learning (Recap) :

**TD Control** ➢ **SARSA** {On-Policy -> State (S), Action (A), Reward (R), State (S’), Action (A’) } ➢ **Q-Learning** {Off-Policy -> Max Q (s,a) }

2). Pengenalan tentang Robotics

• Pengertian (Dasar) dari Robotika

➢ Robotics adalah suatu disiplin ilmu yang mempelajari tentang konsep suatu robot

• Contoh beberapa aplikasi Robot

• Mempelajari Reinforcement Learning for Robotics

Reinforcement Learning for Robotics

➢ Kinematics System

➢ Navigation System

➢ Optimal Control

• Penjelasan State dan Action yang tersedia pada env

Action yang tersedia pada env :

➢ Move left

➢ Move right

• Mempelajari Penerapan DQL pada Cart-Pole

States ➢ Neural Network ➢ Q(s,a)

**PERTEMUAN 7**

Membahas Modul Domain AI Reinforcement Learning Bagian Review and Forum Group Discussion

Materi yang dipelajari hari ini antara lain sebagai berikut :

01 RL Environment

• Mempelajari Dasar-Dasar RL , seperti:

- Penjelasan Konsep RL Algoritme RL menghendaki pertukaran state-action-reward antara agent dan

environment. Teori Dasar, Interaksi Agent, Policy, Value Function, Model

- Membahas MDP (Markov Decision Process)

- Mempelajari Persamaan Bellman Membahas Formulasi dan Perhitungan.

02 Metode-Metode RL

• DP (Dynamic Programming) Membahas Tujuan dari algoritma DP

• MC (Monte Carlo) Membahas Konsep dari Monte Carlo, Elemen Algoritme MC, MC Control, dan

Simple Monte Carlo

• Temporal Difference Learning Mempelajari Konsep Temporal Difference (TD), Perbandingan antara

MC vs TD, Penjelasan tentang Simplest TD Method, dan membahas MC vs TD dalam kasus

sederhana domain RL.

• TD Control: Sarsa & Q-Learning Penjelasan tentang SARSA (State-Action-Reward-State-Action),

Algoritma Sarsa (on- policy TD Control) dan Q-Learning – Off Policy TD Control.

• Deep Q Learning Membahas DQN Deep Q Network dan juga Algoritma DQN. 03 RL untuk Robotics

• Membahas Konsep RL untuk Robotics Mempelajari Contoh beberapa aplikasi Robot : Kinematics

System { Manipulator Robot dan Bipedal Robot} , Navigation System, Optimal Control

• Implementasi RL pada Robotics Mempelajari Q Learning dan Deep Q Learning/ DQN pada Turtlebot.