

Kelompok 2

Selasa, 29 Maret 2022

Reinforcement Learning :

1. Afina Aulia Suandri (Alan Turing)
2. Fairuz Fernanda (Persevere)
3. Wildan Attariq Firmansyah (Better)
4. Irvan Kurniawan (Atlas)
5. Fransiskus Ramaditya Arief Nursanto (Visioner)

Student Activities

1. Cari satu kasus OpenAI yang dapat diselesaikan dengan Monte Carlo

<https://github.com/lukysummer/OpenAI-Monte-Carlo-Control-for-Blackjack>

Program untuk bermain game BlackJack.

Algorithm 11: First-Visit Constant- α (GLIE) MC Control

Input: positive integer $num_episodes$, small positive fraction α , GLIE $\{\epsilon_i\}$

Output: policy π ($\approx \pi_*$ if $num_episodes$ is large enough)

Initialize Q arbitrarily (e.g., $Q(s, a) = 0$ for all $s \in \mathcal{S}$ and $a \in \mathcal{A}(s)$)

for $i \leftarrow 1$ **to** $num_episodes$ **do**

$\epsilon \leftarrow \epsilon_i$

$\pi \leftarrow \epsilon\text{-greedy}(Q)$

 Generate an episode $S_0, A_0, R_1, \dots, S_T$ using π

for $t \leftarrow 0$ **to** $T - 1$ **do**

if (S_t, A_t) is a first visit (with return G_t) **then**

$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha(G_t - Q(S_t, A_t))$

end

end

return π

2. Jelaskan yang anda pahami tentang mekanisme program tersebut!

Mekanismenya adalah, bertujuan untuk mendapatkan Policy dan Q (action-value function table) maksimum dengan cara menerima input episode-episode, membuat Q secara acak, kemudian untuk setiap episode akan membuat policy dengan Q greedy kemudian dengan metode first visit, memperbaiki nilai Q.

Adapun penerapan dari *Reinforcement Learning* ini akan mencapai hasil :

- Dalam estimasi ketika pemain mencapai total 16 untuk kartu di tangan, maka kemungkinan akan berpeluang untuk kalah, mengingat Black Jack membatasi kartu maksimal di tangan pemain maupun *dealer* berjumlah 21.

- Dalam kedua kasus permainan berlanjut, *state-value* meningkat atau sejalan dengan total jumlah kartu di tangan pemain dengan rentang 16 - 21.
- Jika pemain menggunakan kartu As, kartu muka yang dimiliki *dealer* tidak mempengaruhi *state-value* berikutnya.
- Jika pemain tidak memiliki kartu As yang bisa digunakan di kondisi tersebut, nilai dari *state-value* akan berfluktuasi bergantung kepada kartu muka yang dimiliki *dealer*. Contoh pemain dengan rentang kartu 16 - 21 di tangan, maka *state-values* akan lebih tinggi ketika karte *dealer* berada pada rentang 6 - 9.
- *Policy* umum yang terjadi di permainan ini adalah tetap bermain atau berhenti.