Bab III

Metodologi Penelitian

3.1 Pengumpulan Referensi

Referensi awal yang dikumpulkan adalah referensi mengenai penggunaan Reinforcement Learning, terutama algoritma SARSA dan pendekatan MDP. Dari hasil pengumpulan referensi didapat cara untuk menentukan State, Action dan Reward yang akan dipakai.

3.2 Pengumpulan Dataset

Terdapat satu tabel yang menyimpan informasi dari bot dan lawan yang dihadapi oleh bot tersebut. Tabel tersebut menyimpan informasi berupa State, Action serta Reward. Untuk selanjutnya, tabel tersebut akan dibagi menjadi 2 bagian dimana bagian pertama untuk menyimpan State dan bagian kedua akan digunakan untuk menyimpan Action dan Reward.

Bagian Pertama

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ID | Nama | Pos\_x | Pos\_y | HP |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| player1\_x | player1\_y | player2\_x | player2\_y | player3\_x | player3\_y |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| player1\_hp | player2\_hp | player3\_hp | playergen\_hp |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Bagian dari tabel diatas merupakan State yang disimpan, terdiri dari :

-ID : ID record yang telah tersimpan

-Nama : Nama dari bot

-Pos\_x : koordinat posisi bot berdasarkan titik horizontal

-Pos\_y : koordinat posisi bot berdasarkan titik vertikal

-HP : Health Point yang dimiliki oleh bot

-player(1-3)\_x : koordinat posisi lawan bot berdasarkan titik horizontal

-player(1-3)\_y : koordinat posisi lawan bot berdasarkan titik vertikal

-player(1-3)\_hp : Health Point yang dimiliki oleh lawan bot

-playergenhp : Health Point yang dimiliki oleh Generator lawan bot

Bagian Kedua

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| sudut\_x | sudut\_y | Fire | Movement | Reward |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Bagian dari tabel diatas merupakan Action dan Reward yang disimpan, terdiri dari :

-sudut\_x : koordinat sudut tembakan bot berdasarkan titik horizontal

-sudut\_y : koordinat sudut tembakan bot berdasarkan titik vertikal

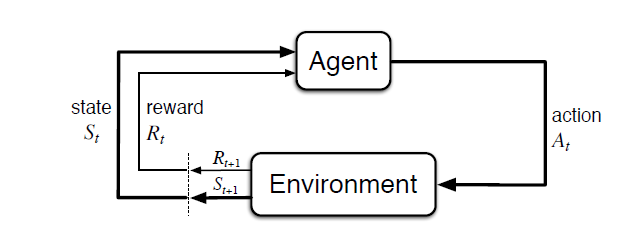
-Fire : menentukan bot menembak atau tidak ( 0 untuk tidak menembak, 1 untuk menembak )

-Movement : Langkah yang diambil oleh bot tersebut

-Reward : Hasil / Poin yang akan diterima oleh bot

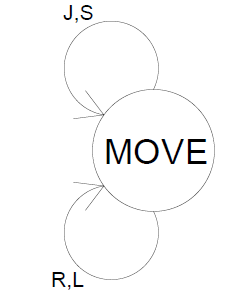
3.3 Analisis

Bot pada Reinforcement Learning ( RL ), terutama algoritma SARSA membuat keputusan ( A ) berdasarkan state (S) yang diterima dari lingkungan berdasarkan posisi bot dari waktu ke waktu.

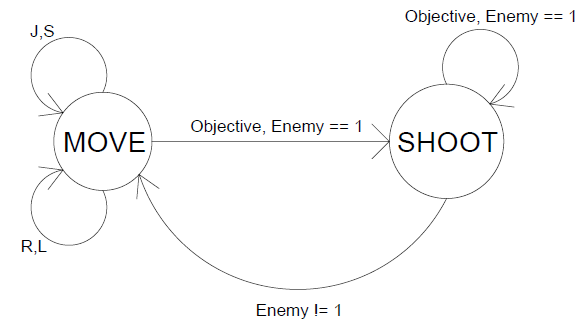


Jika pada state terdapat informasi dari keadaan yang lalu, maka state tersebut dikatakan memiliki Markov property. Dan jika sebuah RL memiliki Markov property, maka RL tersebut adalah Markov Decision Process ( MDP). Permasalahan pada RL seringkali dimodelkan sebagai MDP, dimana S sebagai state, A sebagai Action, P sebagai probabilitas dan R sebagai output atau reward, yang diberikan pada state ( S ), action ( A ) dan state selanjutnya ( S’ ). MDP digunakan untuk memaksimalkan reward dengan memakai fungsi atau policy yang paling optimal berdasarkan lingkungan.

Untuk mengimplementasikan RL pada bot, harus dilakukan sebuah pendekatan dengan MDP. Hal ini penting karena Gang Garrison memiliki sebuah lingkungan yang kompleks dan memerlukan pembagian dari informasi agar didapatkan pembelajaran dan pemilihan keputusan yang optimal.



Untuk pergerakan, bot bergerak dengan berjalan ke kiri (L) , kanan (R), melompat (J) dan berhenti (S). Pada tahap awal pembelajaran, bot akan menyimpan semua koordinat posisi pergerakan dari bot dan menyimpan langkah-langkah yang dipakai oleh bot untuk mencapai koordinat tersebut dan kemudian akan dipilih koordinat terbaik dan akan dilakukan perhitungan reward untuk mengecek optimal atau tidaknya koordinat yang dipilih.



Pada pembelajaran selanjutnya, bot akan dihadapkan dengan adanya musuh dan objective (target utama untuk memenangkan permainan). Bot akan bergerak untuk mencari musuh dan objective, serta menyerang musuh dan objective

Pengambilan keputusan akan dilakukan dengan membaca semua array dalam database, kemudian membandingkan reward dari setiap array untuk mendapatkan reward optimal dan kemudian menjalankan langkah-langkah yang ada pada array yang telah dipilih tersebut. Proses pemilihan keputusan SARSA dengan pendekatan MDP dapat dilihat pada Gambar