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 dua tabel yang menyimpan informasi dari bot pada masa pembelajaran, yakni tabel pergerakan ( GerakRL ) dan tabel menembak ( ), setiap tabel menyimpan State, Action dan Reward secara terpisah.

Tabel pergerakan ( GerakRL )

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Nama | Class | Pos\_x | Pos\_y | Movement | Previous Reward | Reward | TotalReward |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

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

-Nama : Nama dari bot.

-Class : Jenis Karakter yang dipakai oleh bot.

-Pos\_x : koordinat posisi bot berdasarkan titik horizontal.

-Pos\_y : koordinat posisi bot berdasarkan titik vertikal.

-Movement :Langkah yang dipakai oleh bot.

-Previous Reward : Hasil / Poin yang akan diterima oleh bot berdasarkan aksi yang dilakukan sebelumnya.

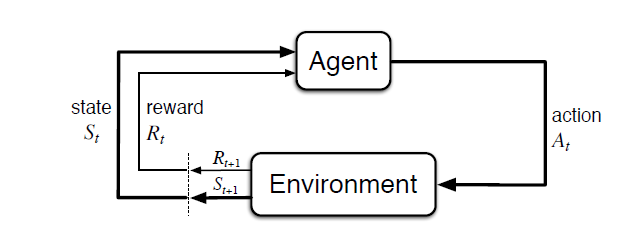
-Reward : Hasil / Poin yang akan diterima oleh bot berdasarkan aksi yang dilakukan.

-Total Reward : Total Reward akhir yang akan diterima oleh bot.

Tabel menembak ()

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 state sebelumnya, maka state tersebut dikatakan memiliki Markov Property. Dan jika sebuah RL memiliki Markov Property, maka RL tersebut dapat disebut sebagai Markov Decision Process ( MDP). Permasalahan pada RL seringkali dimodelkan dalam 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 dimana bot berada.

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.

Langkah – langkah pembelajaran pada bot Reinforcement Learning :

1. Pemberian Reward

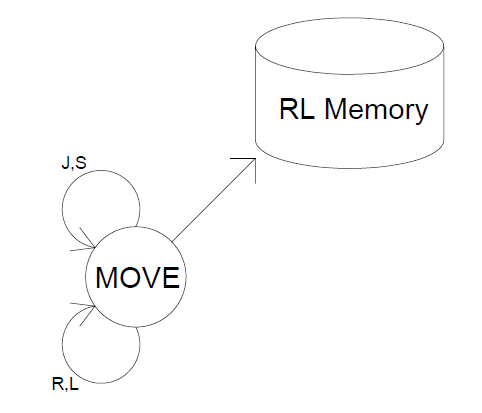
Pemberian Reward pada cara gerak Bot ditentukan dari seberapa jauh bot dapat berpindah dari aksi yang di lakukan.

Total Reward ????

Pemberian Reward pada cara Menembak Bot ditentukan dari ketepatan dalam melakukan tembakan terhadap lawan.

1. Merekam Cara Gerak Bot

Bot akan diberikan instruksi untuk bergerak dengan berjalan ke kiri (L) , kanan (R), melompat (J) dan berhenti (S). Semua koordinat posisi pergerakan dari bot(State) dan langkah-langkah yang diambil oleh bot(Action) untuk mencapai koordinat tertentu(State') akan direkam ke RL Memory.



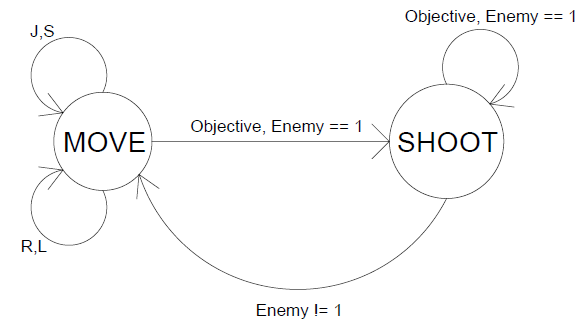
1. Mengolah RL

RL Memory akan dinormalisasi agar pembacaan data akan lebih efisien. Pada penormalisasi gerakan RL Memory dapat dilakukan dengan cara membatasi threshold dari state(S/posisi),dan Action yang bebeda(A) dengan total reward masing masing yang maximum(Rt) serta membuang semua State dan action yang dilakukan dengan reward <= 0. Dalam SQL ( **CREATE VIEW RLGerakbot AS SELECT RL2.nama,printf('%d',RL2.pos\_x) as pos\_x,RL2.movement,RL2.totalreward FROM (SELECT round(pos\_x) as pos\_x,movement,max(totalreward) as totalreward FROM GerakRL --where nama="My Bot 1" group by nama,round(pos\_x),movement) RL1 JOIN (SELECT round(pos\_x) as pos\_x,movement,nama,totalreward FROM GerakRL --where nama="My Bot 1" group by round(pos\_x),movement,nama,totalreward) RL2 ON RL1.pos\_x = RL2.pos\_x where RL1.totalreward=RL2.totalreward and RL1.movement=RL2.movement**).

1. Membaca Data Gerakan pada RL Memory dan Merekam Cara Gerak Bot

Data Gerakan pada RL Memory yang sudah di normalisasi akan dibaca untuk pergerakan bot. Semua koordinat posisi pergerakan dari bot dan langkah-langkah yang diambil oleh bot untuk mencapai koordinat tertentu akan kembali direkam lagi ke RL Memory. Perekaman data tersebut berfungsi sebagai bahan pertimbangan untuk pembelajaran berikutnya dan akan kembali diolah untuk mendapatkan peningkatan dalam pergerakan.





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, setiap Action bot akan disimpan dan kemudian semua Action bot akan dilakukan perhitungan reward untuk mengecek optimal atau tidaknya Action yang dipilih oleh bot.

Pengambilan keputusan dalam pengujian 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