

Jelaskan bagaimana proses dari Q-learning bekerja

Q-learning memiliki tabel untuk mencatat nilai nilai yang akan didapat sesuai kondisi posisi sekarang dan aksi yang akan dilakukan. Pada awalnya tabel ini diinisialisasi dengan nilai 0. Setiap kali agen memilih langkah, indeks yang bersangkutan pada tabel q akan diupdate dengan rumus

```
qTable[currPos][idex] = qTable[currPos][idex] +
lr*(rewards[currPos+action] + gamma*(max(qTable[currPos+action])) -
qTable[currPos][idex])
```

Dengan lr (learning rate) dan gamma adalah untuk menentukan sebanyak apa nilai diupdate setiap terjadi pembaruan nilai q.

Setiap tahapan yang diambil oleh agen adalah antara salah satu dari explore atau exploit. Explore berarti agen menentukan langkah berikutnya secara random untuk mengupdate nilai nilai q dengan harapan mengetahui nilai nilai q dari langkah-langkah baru, exploit adalah bergerak berdasarkan nilai q yang paling tinggi yang sudah diketahui. Q learning memiliki atribut epsilon yang menentukan pada saat itu agen melakukan explore atau exploit, dalam program yang diimplementasikan, jika nilai random kurang dari 0.2, program akan melakukan explore.

Sebutkan kelemahan dari algoritma Q-learning

Kelemahan dari algoritma Q-learning adalah yaitu bergantung pada nilai random, sehingga terdapat kemungkinan loop terjadi ketika program bergerak ke kiri dan kanan secara terus menerus karena hanya nilai di 2 kotak tersebut yang terbarukan sedangkan yang lain masih bernilai 0.