1. Jelaskan bagaimana proses dari Q-learning bekerja
2. Sebutkan kelemahan dari algoritma Q-learning

Jawab

1. Q-learning merupakan algoritma reinforcement learning yang memiliki 2 karakteristik tambahan, yaitu jumlah kemungkinan kondisi (*state*) dan aksi yang terbatas. Q-learning berfokus pada *model-free environment,* artinya AI tidak “belajar” dengan didasari oleh *probability distribution.* Algoritma ini juga menggunakan metode *trial-and-error*, yaitu dengan mencoba menyelesaikan permasalahan dengan berbagai pendekatan dan secara bersamaan memperbarui aturan-aturannya.

Nilai dari setiap kondisi dan aksi dalam permainan akan disimpan ke sebuah Q-value. Q-value disimpan pada sebuah Q-table dengan tiap barisnya merepresentasikan setiap kondisi yang mungkin dan kolom merepresentasikan setiap aksi yang mungkin. Q-table yang optimal akan berisi nilai yang digunakan AI untuk mengambil aksi terbaik untuk mendapat *reward* yang paling tinggi. Q-table inilah yang disebut sebagai aturan (*policy*) yang terus diperbarui.

Q-value sebenarnya berisi estimasi dari jumlah *reward* yang akan diterima. Estimasi ini didapat dari akumulasi dari seluruh langkah yang tersisa pada episode tertentu pada sebuah kondisi dan aksi tertentu. *Reward* ini akan terus bertambah seiring dengan betambah dekatnya si AI kepada tujuannya.

Q-value pada Q-table diperbarui dengan menjumlahkan Q-value sebelumnya dengan hasil kali dari *learning rate* dan *temporal difference*.

Q: Q-value

s: *state*/kondisi

a: aksi

: *learning rate*

TD: *temporal difference*

*Temporal difference* adalah metode yang digunakan untuk mengkalkulasi Q-value yang akan didapat untuk suatu aksi pada suatu kondisi berdasarkan apa yang telah dipelajari oleh AI mengenai Q-value untuk aksi di kondisi terkini.

r: *reward*

*: discoumt factor*

Setelah Q-table telah sepenuhnya diperbarui, maka AI dapat dikatakan telah sepenuhnya dilatih. Langkah selanjutnya adalah mode inferens, yaitu AI akan memilih aksi terbaik berdasarkan Q-value yang diketahui.

1. Kelemahan q-learning
   1. Perlu menentukan *learning rate* dan *discount factor* secara manual
   2. Semakin banyak *state* dan aksi, semakin banyak *space* yang dibutuhkan untuk menyimpan Q-table