

Q-LEARNING

a. Jelaskan bagaimana proses dari Q-Learning bekerja?

Secara ringkas q-learning terbagi ke dua komponen utama yakni:

1. Reward and punishment
Skor atau nilai yang diberikan ketika melakukan aksi tertentu di keadaan tertentu
2. Exploration and exploitation
Exploration : mencoba melakukan suatu aksi pada suatu keadaan tertentu
Exploitation : maksimasi reward dengan melakukan aksi yang memberi skor terbesar

Selain itu terdapat beberapa hyperparameter yaitu sebagai berikut:

1. Jumlah episode
Rangkaian kejadian-kejadian yang dilakukan oleh agen di suatu environment
2. Jumlah langkah per episode
Jumlah langkah maksimum agen pada satu episode
3. Learning rate
Kecepatan belajar agen
4. Discount rate
Pengali dari reward maksimum sebelumnya dari kejadian yang sudah pernah dilakukan
5. Exploration decay rate
Laju untuk kapan agen melakukan observasi

Penjelasan:

Dalam q learning, akan dibuat sebuah tabel dengan dimensi m kali n dimana m merupakan banyaknya keadaan dan n merupakan banyaknya aksi yang bisa dilakukan. Tabel tersebut dinamakan q table dan diinisiasi semuanya dengan nilai 0. Agen akan diberikan perintah untuk beraktivitas di sebuah environment (dalam kasus ini game) sampai dengan gamenya usai atau agen telah sampai kepada batas langkah yang bisa dilakukan, ini dinamakan satu episode. Setiap aksi yang dilakukan oleh agen akan menyebabkan agen memperoleh sebuah angka yang merepresentasikan imbalan atas aksinya di suatu keadaan tertentu, ketika aksi sang agen menuju kemenangan maka agen diberi reward pada sel di tabel pada baris keadaan yang terkait dan kolom aksi yang terkait, begitu pula jika agen melakukan kesalahan maka agen diberi punishment dan angka-angka yang diperoleh tersebut akan dikalkulasikan.

Setelah selang beberapa episode dan langkah, maka agen memiliki data perolehan skor pada masing-masing sel (aksi yang dilakukan pada kondisi tertentu). Harapannya adalah agen bisa maksimisasi nilai yang diperoleh sehingga agen akan memilih aksi-aksi yang menguntungkannya, ini dinamakan exploitation. Namun, ada beberapa keadaan dimana ternyata agen tidak tahu ada aksi lain yang ternyata memberikan reward yang jauh lebih besar dibandingkan aksi yang selama ini dieksploitasi, maka dari situ agen diberikan kesempatan untuk eksplorasi (observasi lingkungan sekitarnya).

Setelah berepisode-episode maka agen akan menemukan caranya untuk sampai kepada kemenangan.

a. Sebutkan kelemahan dari algoritma Q-Learning

Kelemahan dari q learning adalah:

- Komputasi yang mahal karena dimensi dari tabel yang luas
- Skema pemberian reward dan punishment harus diperhatikan secara detail karena akan mempengaruhi keberjalanan agen
- Pemberian nilai hyper parameter yang harus lebih diperhatikan juga