

1. Deskripsi Masalah

Diberikan sebuah data berisi 225 angka dengan format matriks 15x15, matriks tersebut adalah *reward* dari sebuah perjalanan, dimana perjalanan tersebut kita harus mendapatkan *reward* yang besar, maka berikut adalah algoritmanya :

1. Set Qtable, dan set nilai tak hingga apabila tidak bisa melakukan pergerakan, misal di ujung atas, maka tidak bisa bergerak ke atas, Q table berisi 225 kolom (dari total 15x15) dan 4 baris (4 arah gerak)

Qtab - NumPy array

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
0	-inf	-inf	-inf	-inf	-inf	-inf	-inf	-inf	-inf	-inf	-inf	-inf	-inf	-inf	-inf	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-inf	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	-inf	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-inf	0	0	0	0	0	0	0

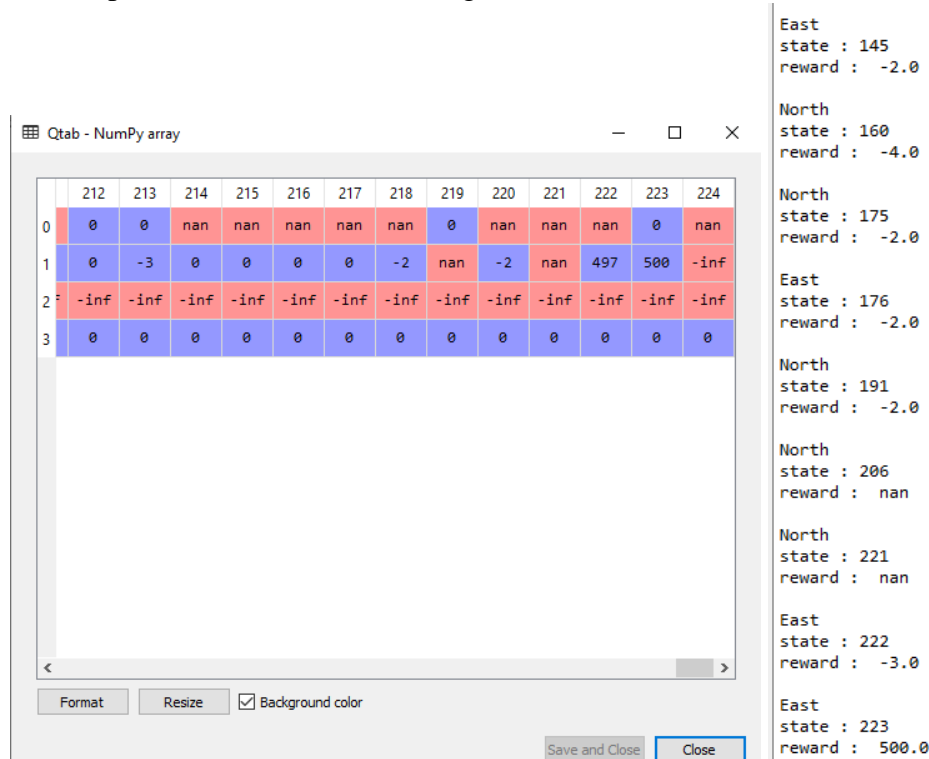
2. Karena ada total 225 data, maka untuk mengambil nilai reward dibutuhkan sebuah array 15x15 berisi urusan indeks dari state (poin x,y)0-224
3. Set gamma =1 dan alpha =1
4. Melakukan pencarian nilai *reward* maksimum dengan mengupdate Qtable
 - a. Melakukan random untuk menentukan step
 - b. Bila random = 0 maka state berkurang 15 (karena north berarti naik ke atas dan mengundangi 15 langkah ke belakang karena state 0 berada di bawah)
 - c. Bila random= 1 maka state bertambah 1 (karena east, yang berarti bergerak ke kanan dan menambah state 1)
 - d. Bila random =2 maka state bertambah 15 (karena south berarti turun ke bawah dan menambah 15 langkah ke depan karena state 224 berada di atas)
 - e. Bila random =3 maka state berkurang 1 (karena west, bergerak ke kiri dan state berkurang 1)
5. Update Q table dengan persamaan :
 - a. $Qtab[arah][state] = Qtab[arah][state] + \alpha * (reward[rd][sacc] + (\gamma * (\max(Qtab[kolom\ state+1])) - Qtab[arah][state]))$
6. State = state+1

1. Analisis

Pada permasalahan kali ini terdapat sejumlah 225 data dimana berbentuk array dan dimulai pada kotak pertama di titik array dengan row = 14 column = 0 dan berakhir di titik array dengan row = 0 dan column = 14 dengan reward 500. Kita diminta untuk mendapatkan *reward* sebesar besarnya.

Berikut adalah hasil runningnya :

Hasil Update Qtab dan hasil running



Didapatkan hasil running terbaik dengan hasil *reward* = 500, tidak dilakukan perulangan sebanyak n jumlah episode karena dengan algoritma ini, state akan menentukan sendiri jalan dan berhenti saat reward telah menyentuh nilai maksimum

Dalam program ini bisa terjadi looping forever, diduga terjadi karena hasil dari tabel bernilai nan, dan tidak terdeteksi nilai maksimumnya, double scalar error juga terjadi diduga karena mengalikan nilai dengan 0 atau nan, namun jika dilakukan running kembali, bisa menjadi normal lagi.