LAPORAN TUGAS 3

REINFORCEMENT LEARNING: Q-LEARNING

Nama : Adhyfa Fahmy Hidayat

Kelas : IF 39-01

NIM : 1301154127

Analisis Masalah

Dalam tugas 3 ini, kita harus membuat sebuah sistem Q-Learning untuk menemukan optimum policy dari posisi start (kuning) sampai ke posisi goal (hijau). Agent hanya bisa melakukan 4 aksi: atas, bawah, kiri, kanan.

10	-1	-3	-5	-1	-3	-3	-5	-5	-1	100
9	-2	-1	-1	-4	-2	-5	-3	-5	-5	-5
8	-3	-4	-4	-1	-3	-5	-5	-4	-3	-5
7	-3	-5	-2	-5	-1	-4	-5	-1	-3	-4
6	-4	-3	-3	-2	-1	-1	-1	-4	-3	-4
5	-4	-2	-5	-2	-4	-5	-1	-2	-2	-4
4	-4	-3	-2	-3	-1	-3	-4	-3	-1	-3
3	-4	-2	-5	-4	-1	-4	-5	-5	-2	-4
2	-2	-1	-1	-4	-1	-3	-5	-1	-4	-1
1	-5	-3	-1	-2	-4	-3	-5	-2	-2	-2
	1	2	3	4	5	6	7	8	9	10

Figure 1: Sebuah *grid world* ukuran 10 x 10, di mana angka-angka dalam kotak menyatakan *reward*. *Agent* berada di posisi *Start* (1,1) dan *Goal* di posisi (10,10)

Yang harus dicari adalah si agent lewat mana saja sampai ke goals di posisi matriks (10,10)

Desain

Algoritma Q-Learning secara umum adalah sebagai berikut:

- Tentukan parameter gamma
- Tentukan reward dalam matriks R
- Inisialisasi matriks Q ke 0
- Perulangan for:

  per episode sampai goal

  pilih random state sebagai current state (initial)

perulangan while:

sampai goal belum tercapai

- Pilih action yang memungkinkan
- Lalu lanjut bergerak ke action yang memungkinkan tersebut
- Cari nilai Q yang paling maksimum berdasarkan seluruh action yang memungkinkan
- Hitung nilai Q menggunakan rumus yang sudah ditentukan

$$Q(state, action) = R(state, action) + Gamma * Max[Q(next\ state, all\ actions)]$$

- Tentukan next state sebagai current state

End while

End for

Hasil Eksperimen

Pada file reward.xlsx berisi reward yang didapat dari table matriks DataTugasML3.txt pada setiap possible actionnya. Dapat dilihat bahwa goal berada di baris ke 91 dan initial state pada baris ke 10.

Berikut adalah Q akhir yang sudah dihitung dengan gamma = 0,8

Q =

0	2.9294	0	1.9294
3.9294	1.9294	0	1.3435
2.9294	0.3294	0	-0.6706
-0.6565	-6.1892	0	-5.6264
-3.5252	-5.5605	0	-0.7830
-6.8202	-8.3411	0	-2.6264
-6.9892	-8.3411	0	-2.6417
-8.4484	-5.1684	0	-4.1133
-8.1347	-10.5212	0	-4.2907
-7.5212	0	0	-6.4325
0	1.3435	6.1617	-1.8565
1.9294	-2.9252	2.9294	0.0748
1.3435	-5.6264	0.3294	-0.8565
-2.9252	-0.7830	-3.5252	-2.6852
-4.2565	-2.6264	2.7713	-3.6005
-0.7830	-2.6417	-6.9892	-4.6417
-2.6264	-4.1133	-8.3411	-4.1133
-2.6417	-4.2907	-7.1684	-6.9856
-4.1133	-6.4325	-7.5212	-2.9856
-4.2907	0	-9.2522	-3.3885
0	3.9294	1.9294	-0.7432
0.3210	-0.8565	1.3435	-0.8565
3.9294	-2.6852	-2.9252	-1.6852
-0.8565	-2.7557	-5.6264	-3.6701
-0.7506	-4.6417	-0.7830	1.6624
-3.6005	-4.1133	0.4479	-0.6701
-4.6417	-7.3885	-2.6417	0.6624

-4.6417	-7.3885	-2.6417	0.6624
-4.1133	-2.9856	-4.1133	-3.4701
-7.3885	-2.4820	-4.2907	-5.9723
-2.9856	0	-4.4820	-4.7108
0	-2.0345	0.3210	-2.3566
-0.3566	-1.6852	2.4568	-1.7890
-2.0345	-3.6701	-0.8565	-2.3396
-1.6852	1.6624	-2.5746	0.8255
-3.6701	4.5780	-3.6005	0.3299
1.6624	0.6624	-4.6417	1.1024
4.5780	-3.4701	-4.1133	-0.1181
0.6624	-5.9723	-7.3885	-1.0945
-3.4701	-3.9723	-2.9856	-0.5818
-6.3885	0	-3.3885	-3.5818
0	-1.7890	-0.7432	-2.3566
-3.5945	-2.3396	0.2638	-6.4312
-1.7890	0.8255	-1.6852	-6.8717
-2.3396	2.2819	-3.6701	-3.3396
0.8255	-2.1745	1.6624	0.8255
2.2819	-0.1181	-0.6701	-4.1181
-2.1745	-1.0945	0.6624	-1.0965
-0.1181	-0.5818	-3.4701	-3.5204
-1.0945	-2.7204	-5.9856	-3.2319
-0.5818	0	-1.5818	-2.5818

-0.5818	0	-1.5818	-2.5818
0	-6.4312	-3.5945	-7.8756
-5.8756	-6.8717	-1.7890	-8.1450
-6.4196	-3.3396	-2.3396	-6.8381
-6.8717	0.8255	0.8255	-5.2717
-3.3396	-4.1181	2.2819	-0.3396
0.8255	-1.0965	1.1024	0.8255
-1.7181	-4.8756	-0.1181	-3.3396
-1.0965	-3.2319	-1.0945	-7.6717
-4.8756	-3.2372	-0.2899	-4.5204
-3.2319	0	-2.7204	-2.2965
0	-8.1357	-5.8756	-9.4519
-9.7005	-6.8381	-6.4312	-9.9533
-7.8513	-3.1745	-6.8717	-6.1917
-7.5396	-0.3396	-3.3396	-2.7396
-2.2976	0.8255	0.8255	-2.1745
2.2819	-3.3396	-4.1181	1.3016
0.8255	-7.6717	-1.0965	-4.0717
-3.3396	-4.5204	-4.8756	-4.9172
-7.6717	-6.8372	-3.2319	-2.5338
-4.5204	0	-2.9388	1.1945
0	-9.9533	-9.7005	3.2995
-12.5587	-6.1917	-8.1450	-2.3604
-6.0641	-2.7396	-6.8381	-1.9005

```
   -6.0641   -2.7396   -6.8381   -1.9005
   -6.1917   -2.1745   -5.2717   -0.9729
   -2.7396   -1.3396    2.2819   -1.6983
   -2.1745    0.5823    0.8255    1.6271
   -1.3396   -4.5342   -3.3396    4.5338
   -4.0717   -2.4444   -3.0965    1.8900
   -6.9555   -1.9172   -4.5204    4.8625
   -2.4444         0   -0.4772    6.8625
         0   -3.6884   -0.7005  499.9943
    1.6396   -3.5604   -6.0641  394.9963
   -0.7005   -0.9729   -6.1917  310.9963
   -3.5604   -1.6983   -2.7396  244.7971
   -4.3587    1.6271   -2.1745  191.8363
   -1.6983    4.5338   -1.3396  149.4683
    1.6271    6.9173   -4.0717  116.5730
    4.5338    4.8625   -4.5342   89.2592
    1.8900   11.0781   -2.5338   70.4067
    4.8625         0   11.0781   54.3254
         0  394.9963  398.9963         0
  499.9963  310.9963   -2.3604         0
  394.9954  244.7971   -3.5604         0
  310.9963  191.8377   -0.9729         0
  244.7971  149.4683   -1.6983         0
191.8354  116.5739    1.6271         0
149.4683   89.2584    4.5338         0
116.5739   70.4067    1.8900         0
 89.2584   54.3254    4.8625         0
```

Setelah mencapai current 91 yang dimana merupakan goal state maka perulangan berhenti. Setelah itu mendapatkan total reward dari initial state sampai goal state sebanyak:

```
total_reward =

  495.6065
```