

TUGAS 2 Machine Learning

Q-Learning



Disusun oleh

Dzulfiqar Ridha

1301154298

IF-39-04

Telkom University

Jl. Telekomunikasi No. 01, Terusan Buah Batu, Sukapura, Dayeuhkolot, Sukapura,
Dayeuhkolot, Bandung, Jawa Barat 40257

2018

1. Deskripsi Masalah

Bangunlah sebuah sistem Q-learning untuk menemukan optimum policy sehingga Agent yang berada di posisi Start (1,1) mampu menemukan Goal yang berada di posisi (10,10) dengan mendapatkan Total Reward maksimum pada grid world dalam Figure 1 berikut ini. Data pada Figure 1 dapat dilihat di file DataTugasML3.txt. Pada kasus ini, Agent hanya bisa melakukan empat aksi: N, E, S, dan W yang secara berurutan menyatakan North (ke atas), East (ke kanan), South (ke bawah), dan West (ke kiri). Anda boleh menggunakan skema apapun dalam mengimplementasikan sebuah episode.

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

GambarError! No text of specified style in document..1 Reward Tabel (kuning start, hijau goal)

2. Detail Rancangan Metode yang Digunakan

Analisa Masalah Akan dibangun sebuah sistem Q-Learning untuk menemukan optimum policy sehingga Agent yang berada di posisi Start(9,0) mampu menemukan Goal state yang berada di posisi (0,9) dengan mendapatkan Total Reward maksimum pada grid world dalam Figure 1 berikut ini. Data pada Gambar 1 dapat dilihat di _le DataTugasML3.txt. Pada kasus ini, Agent hanya bisa melakukan empat aksi: atas, kanan, bawah dan kiri. Anda boleh menggunakan skema apapun dalam mengimplementasikan sebuah episode, untuk mengisi reward pada matrix Q. Note: jika pada array matriks [0][0] sama dengan (1,1) sehingga (9,9) = (10,10).

```
[[-1.0, -3.0, -5.0, -1.0, -3.0, -3.0, -5.0, -5.0, -1.0, 100.0],
 [-2.0, -1.0, -1.0, -4.0, -2.0, -5.0, -3.0, -5.0, -5.0, -5.0],
 [-3.0, -4.0, -4.0, -1.0, -3.0, -5.0, -5.0, -4.0, -3.0, -5.0],
 [-3.0, -5.0, -2.0, -5.0, -1.0, -4.0, -5.0, -1.0, -3.0, -4.0],
 [-4.0, -3.0, -3.0, -2.0, -1.0, -1.0, -1.0, -4.0, -3.0, -4.0],
 [-4.0, -2.0, -5.0, -2.0, -4.0, -5.0, -1.0, -2.0, -2.0, -4.0],
 [-4.0, -3.0, -2.0, -3.0, -1.0, -3.0, -4.0, -3.0, -1.0, -3.0],
 [-4.0, -2.0, -5.0, -4.0, -1.0, -4.0, -5.0, -5.0, -2.0, -4.0],
 [-2.0, -1.0, -1.0, -4.0, -1.0, -3.0, -5.0, -1.0, -4.0, -1.0],
 [-5.0, -3.0, -1.0, -2.0, -4.0, -3.0, -5.0, -2.0, -2.0, -2.0]]
```

GambarError! No text of specified style in document...2 Reward table pada implementasi code

	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	...	9.0	
N	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	...	-4973.109792	-1754.61
E	-2542.984494	-1082.422541	-1070.316544	-3165.180160	-1107.675275	-2700.033556	-2019.363115	-774.413261	50300.000000	0.0	...	-4976.839951	-1754.81
W	0.000000	-1076.453809	-1073.985101	-3167.787423	-1105.987743	-2701.853867	-2013.610091	-780.221224	50074.377146	0.0	...	0.000000	-1759.81
S	-2539.655295	-1076.212608	-1071.285808	-3163.567175	-1106.781494	-2704.110853	-2017.285013	-775.282821	50090.951652	0.0	...	0.000000	0.01

4 rows x 100 columns

Gambar 3 Q Tabel pada implementasi code

3. Hasil

Setelah melalui tahapan yang ada, implementasi q-learning menghasilkan hasil seperti pada gambar berikut:

```
selesai dengan reward: 46.0
path yang dilalui: [9.0, 9.1, 9.2, 9.3, 9.4, 9.5, 9.6, 9.7, 9.8, 9.9, 8.9, 7.9, 6.9, 5.9, 4.9, 3.9, 2.9, 1.9, 0.9]
```