Nama : Edy Santoso Kelas: IF 40-08 Nim : 1301162111

Laporan Mencari Optimum Policy dengan Metode Q-Learning

DESKRIPSI MASALAH

Diberikan sebuah word grid seperti dibawah ini.

-1	-2	-3	-2	-3	-3	-4	-1	-4	-2	-1	-2	-3	-3	500
-1	-3	-1	-2	-4	-1	-4	-1	-4	-2	-4	-2	-2	-2	-1
-4	-2	-1	-4	-2	-1	-2	-4	-2	-3	-2	-1	-2	-4	-4
-4	-2	-4	-1	-3	-2	-3	-2	-4	-2	-4	-1	-2	-4	-2
-4	-2	-2	-3	-2	-3	-1	-1	-4	-2	-1	-3	-4	-2	-4
-4	-3	-3	-4	-2	-3	-4	-2	-2	-1	-1	-2	-1	-2	-1
-2	-3	-2	-1	-1	-3	-2	-1	-4	-3	-1	-1	-2	-3	-3
-3	-1	-1	-4	-4	-3	-1	-2	-3	-1	-1	-4	-4	-3	-3
-3	-1	-4	-2	-3	-3	-1	-4	-4	-4	-2	-2	-2	-2	-1
-3	-4	-4	-2	-3	-4	-3	-3	-2	-2	-3	-4	-3	-4	-1
-3	-4	-1	-1	-1	-4	-4	-4	-4	-1	-2	-4	-2	-2	-1
-1	-3	-3	-3	-3	-3	-3	-3	-4	-1	-2	-4	-1	-2	-4
-2	-2	-1	-2	-2	-2	-4	-3	-1	-4	-1	-4	-2	-2	-2
-2	-1	-3	-1	-4	-4	-1	-3	-3	-1	-1	-2	-3	-4	-3
-2	-2	-1	-4	-4	-4	-2	-2	-3	-1	-2	-2	-1	-1	-3

Pada word grid diatas, misalkan terdapat seorang agent yang terdapat pada titik(1,1) dan akan berjalan hingga menjacapi titik (15,15) yang nantinya akan mendapatkan reward maksimum. Setap perjalanan/grid akan mendapatkan reward yang berbeda beda. Maka dari itu diperlukan sebuah sistem untuk mencari jalan hingga sampai di titik (15,15) dengan mendapatkan reward yang maksimum.

METODE PENYELESAIAN

Pembuatan sistem untuk permasalah diatas adalah dengan menggunakan Q-learning. Q-Learning merupakan salah satu algortima yang biasanya digunkan untuk pencarian jalur.

1. Membaca file yang telah diberikan.
2. Menginisialisasi table Q(s,a). tabel dibuat dengan ilustrasi sebagai berikut.

	N	**E**	**S**	**W**
0	0	0	0	0
1	0	0	0	0
…	0	0	0	0
224	0	0	0	0

Pada tabel diatas, terdapat 1 sampai 225 yang mana menandakan banyaknya jalan untuk eksplorasi yang dapat dilakukan yaitu 225 langkah. Sedangkan pada tanda N mempunyai arti north yaitu jalan keatas. S mempunyai arti south yaitu jalan kebawah, W mempunyai arti west mempunyai arti kiri, dan E mempunyai arti east yaitu kekanan.

3. setelah itu, lakukan eksplorasi. Tujuan dalam melakukan eksplorasi adalah untuk mencari jalan hingga mencapai tujuan yang akan diinginkan. pada tahap ini, eksplorasi dilakukan sebanyak 3000 episode dimana setiap episodenya terdiri dari 100 step. Disetiap stepnya akan mengupdate tabel QSA yang telah dibuat apda tahap kedua. Dalam perubahan QSA, terdapat rumus atau persamaan yang dapat dilakukan untuk mengupdate tabel QSA. Adapuny rumus QSA adlah sebagai berikut:

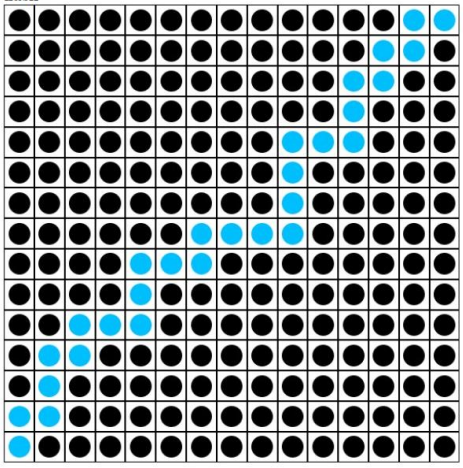
$$Q(s, a) \square Q(s, a) + \alpha[r + \gamma \max_{a'} Q(s', a') - Q(s, a)]$$

Pada rumus diatas, α adalah alpha, yang mana kita dapat menentukan besarnya. Kemudiana r adalah reward, $_{\gamma}$ max adalah maksimum pada tabel Q(s,a) pada setiap arahnya.

4. setelah melakukan ekplorasi, maka langkah selanjutnya dalah ekploitasi. Pada tahap ini, tabel Q(s,a) telah mencapai tahap final. Kemudian. Lihat disetiap langkahnya, arah mana yang memungkinkan untuk menuju ke langkah selanjutnya yang diambil berdasarkan nilai Q(s,a) terbesar disetiap arahnya.

HASIL OUTPUT

hasil



REWARD : 448

tabel qsa

0	0	-33.47374328772747	-33.20485020506172	0
1	0	-15.532442313034307	-15.710949526225988	-15.379703382430426
2	0	-7.827652708281159	-7.985454235588177	-7.730349216119959
3	0	-9.324131732969796	-9.574071256108288	-9.324131732969796
4	0	-4.460061539032589	-4.845892764580989	-4.600773789147327
5	0	10.39491083128521	-3.0851458800791485	-3.0202901051861515
6	0	117.99488640953624	-1.2619260976578204	-1.124822035593291
7	0	581.9262336864085	-2.9596250813281304	-2.91409829388877
8	0	2406.8387677125947	-0.4105312710510279	-0.4025769861695721
9	0	8782.402758548804	-0.26338759913836735	-0.28854140842804
10	0	29384.973285485885	-0.31967690008778	-0.2989484264752446
11	0	-0.09090299999999998	91494.15975172691	-0.0936852726843609
12	0	-0.03	-0.02	-0.02
13	0	449631243586.56384	0	0
14	0	0	0	0
15	-46.02963995000209	-45.8275661509727	-45.586968480930466	0
16	-2.7532374998150626	-2.9988404497168903	-2.814286363867252	-2.7189585619251266
17	-2.3961288076161096	-2.449557428645413	-2.4275480117222337	-2.3946396013198528
18	-1.856920306112185	-2.0153630391015223	-2.3087917568681067	-1.9926398473830058
19	-1.5563696839000323	-1.513832241715201	-1.8997068364006235	-1.516146752700464
20	-1.3352294142824888	-1.4822019472296322	-1.597207935132659	-1.350219003469689
21	4.238970268002263	-0.8771133677635966	-0.8123119641946823	-0.9630082351818321
22	-0.711410468784349	-0.7743759766697095	-0.7804365202320519	-0.8050611859794158
23	-0.3314268225123204	0.046449521906087365	-0.3137439916136971	-0.3662099423935997
24	-0.2092442508224090 2	8.25591694318146	-0.24409007989357026	-0.2683366219251412
25	-0.07213535210701	269.7853563178541	-0.07666766	-0.09516657010518573

Gambar diatas adalah hasil output dari program. Pada gambar disebelah kanan, terdapat tabel Q(s,a) namun tidak semua ditampilakn. Pada gambar ini menampilakn bagaimana kondisi tabel Q(s,a) pada episode terakhir. Pada gambar disebelah kiri adalah gambaran bagimana perjalanan yang diambil hingga mencapai tujuan akhir. Pada langkah yang diambil reward yang didapatkan adlah 448.