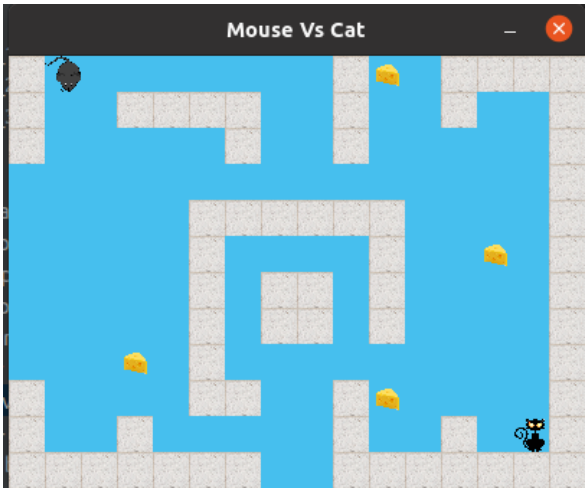
Reprezentarea: pentru reprentarea am creat clasa Map care retine
date citite din fisierul de congfigurarea si are si fuctii ajutatorea
pentru algoritmul q-learning si sarsa, ea se gaseste in fiserul Map. Pentru afisarea jocului am folosit pygame.
(clasa Display din src/display)Cum pentru exploare era la alegrea noastra, am implementat trei startegi: se alege cele
care au fost vizitate cel mai putin, se foloseste o varinta modificat de uct, de la MCTS, alegandu-se maximul sau se
alege probabilistic, calculand probabilitatile folosind softmax.
Pentru exploare/exploate am doua strategi: aleg cu o probabilitate de epsilor intre random/max_first; aleg actionea
probabilistic, probalitatile sunt calculate cu ajutorul functie softmax.

Reprezenaterea celor trei harti folosite la testare:

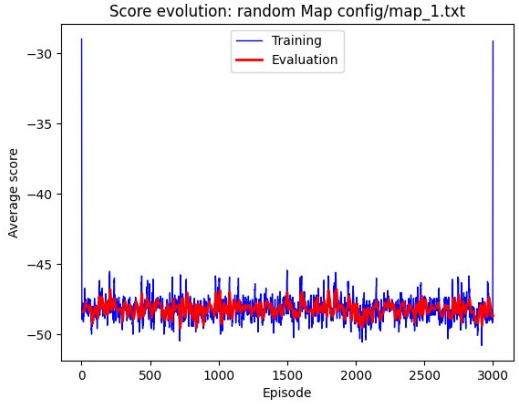
Map 1:

Numar de rulari : 3000

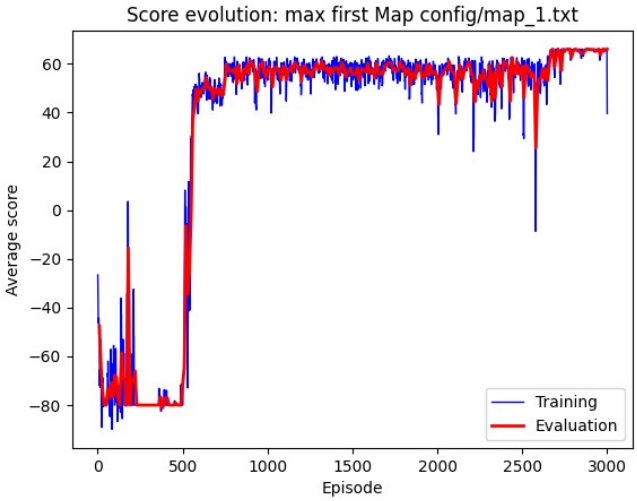


I) Evolutia scorului în funcție de numărul episodului de antrenament:

1) Random:

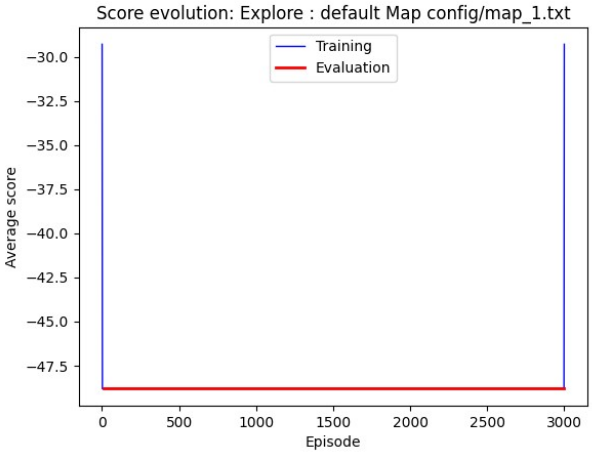


2) Max First:

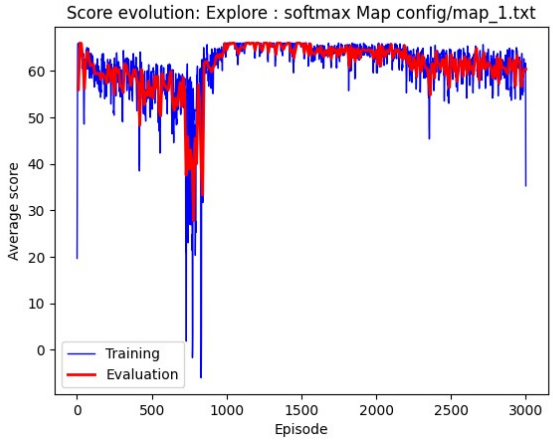


3) Exploatare(dar descrierea pt exploare):

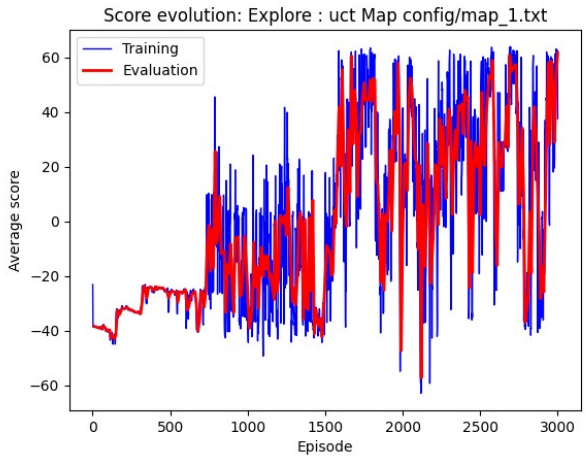
a) default:



b) softmax:

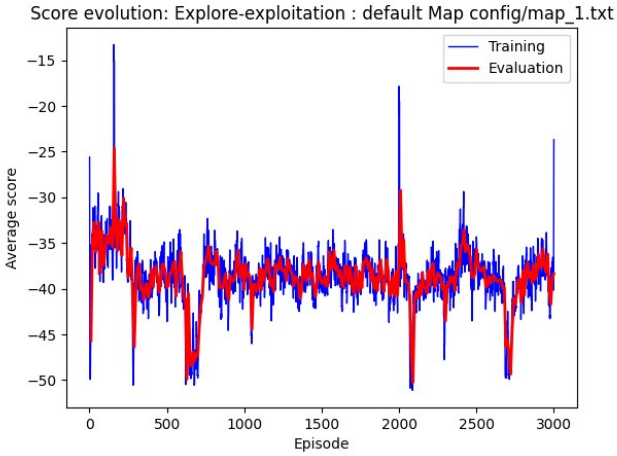


c) uct:

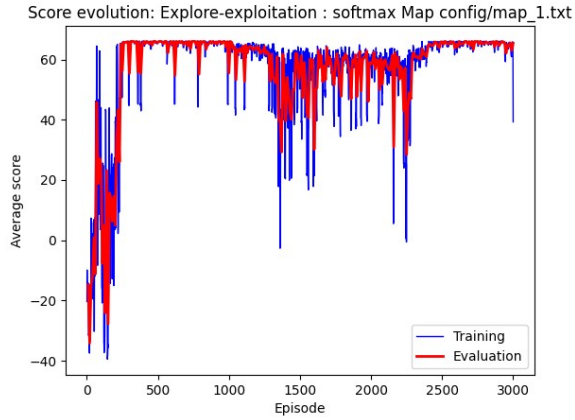


4) Eploare/Exploatare:

a) default:



b) softmax



II) Procentul de jocuri castigate în funcție de valoarea:

1) Random

B	C	D	E	F	G	H	I
D.F. / L.R	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.9	0	0	0	0	0	0	0
0.85	0	0	0	0	0	0	0
0.8	0	0	0	0	0	0	0
0.75	0	0	0	0	0	0	0
0.7	0	0	0	0	0	0	0
0.65	0	0	0	0	0	0	0
0.5	0	0	0	0	0	0	0

2) Max First:

D.F. / L.R	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.9	0.8537	0.9187	0.9283	0.95	0.9487	0.922	0.9427
0.85	0.9193	0.0017	0	0	0.942	0	0.9373
0.8	0	0.927	0.8757	0.8333	0	0.001	0.9327
0.75	0.932	0.0027	0.002	0.946	0.6437	0.9243	0.0003
0.7	0.9007	0.2547	0.9287	0.025	0.948	0.0473	0.949
0.65	0.931	0	0.9427	0	0.939	0.9517	0
0.5	0.0083	0.017	0.947	0.9153	0.1737	0.9397	0.9353

3) Explore:

a) default:

D.F. / L.R	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.9	0	0	0	0	0	0	0
0.85	0	0	0	0	0	0	0
0.8	0	0	0	0	0	0	0
0.75	0	0	0	0	0	0	0
0.7	0	0	0	0	0	0	0
0.65	0	0	0	0	0	0	0
0.5	0	0	0	0	0	0	0

b) softmax:

D.F. / L.R	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.9	0.349	0.3373	0.9043	0.7933	0.218	0.3643	0.2613
0.85	0.3667	0.2903	0.3533	0.3877	0.2757	0.5	0.314
0.8	0.3513	0.3967	0.31	0.374	0.2053	0.489	0.2863
0.75	0.3493	0.3713	0.362	0.2457	0.418	0.343	0.2067
0.7	0.3457	0.359	0.4917	0.312	0.3543	0.254	0.199
0.65	0.343	0.4767	0.3997	0.2967	0.311	0.2473	0.2607
0.5	0.29	0.3057	0.3213	0.261	0.2553	0.2243	0.251

4) Eploatare/Explorare:

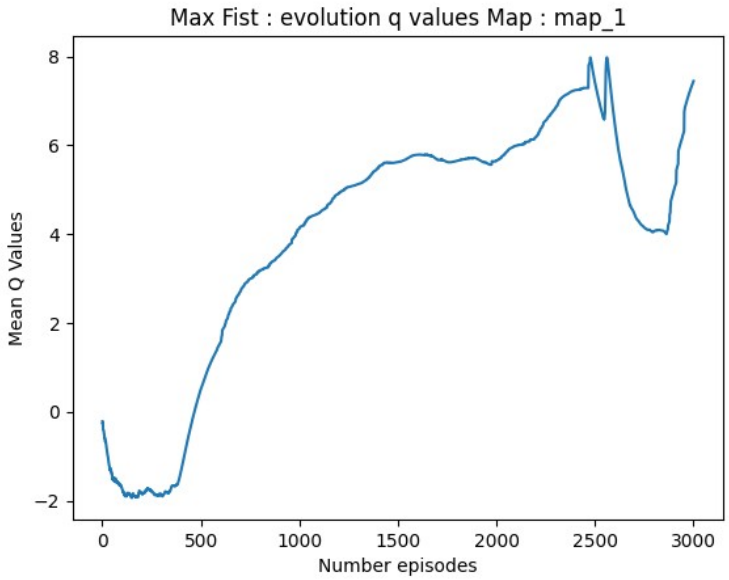
a) default:

D.F. / L.R	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.9	0	0	0	0	0	0	0
0.85	0	0	0	0	0	0	0
0.8	0.0003	0	0	0	0	0	0
0.75	0	0	0	0	0	0	0
0.7	0	0	0	0	0	0	0
0.65	0	0	0	0	0	0	0
0.5	0	0	0	0	0	0	0

b) softmax:

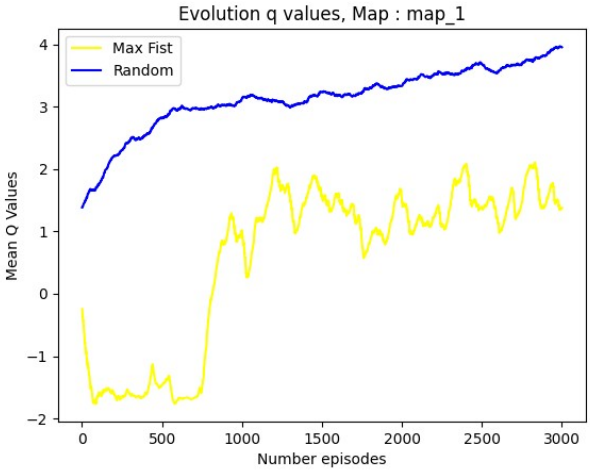
D.F. / L.R	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.9	0.728	0.8377	0.2603	0.2557	0.687	0.221	0.0763
0.85	0.3213	0.3357	0.471	0.6403	0.183	0	0.257
0.8	0.374	0.3783	0.624	0.3003	0.235	0.4583	0.1883
0.75	0.3607	0.323	0.3993	0.2883	0.373	0.311	0.3113
0.7	0.3547	0.3897	0.3713	0.2143	0.6343	0.2963	0.194
0.65	0.3423	0.3323	0.2997	0.3303	0.259	0.0097	0.2523
0.5	0.3043	0.307	0.267	0.2873	0.351	0.3013	0.211

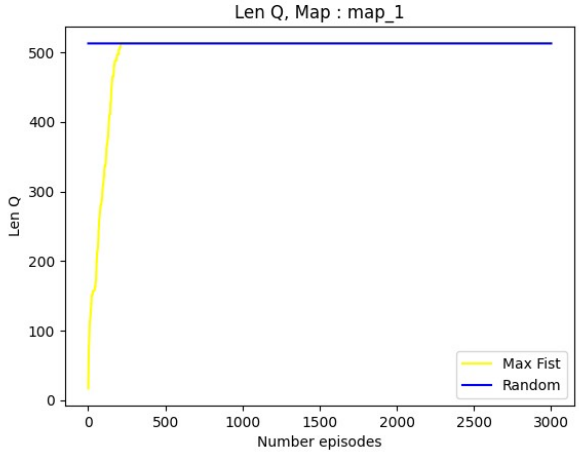
III) Cum afectează numărul de episoade de antrenament valorile din tabela de utilitati în cazul strategiei max first?



Se poate observa ca are o scade pana cand gaseste cale/drumul optim catre casting.

IV)Care sunt diferențele intre tabela de utilitati din cazul strategiei max first și random?

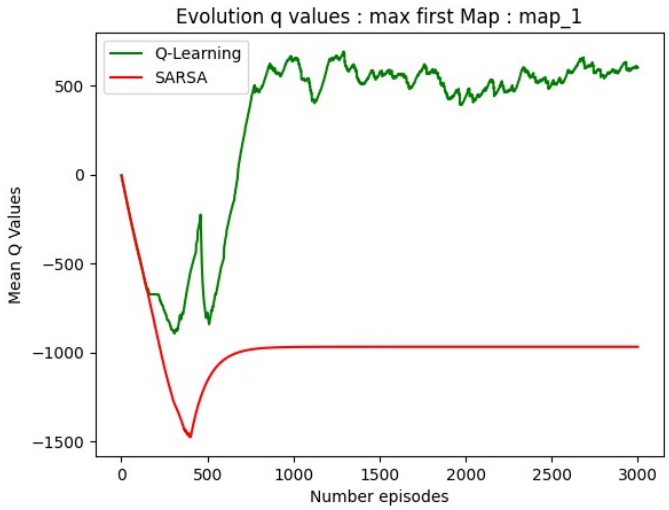


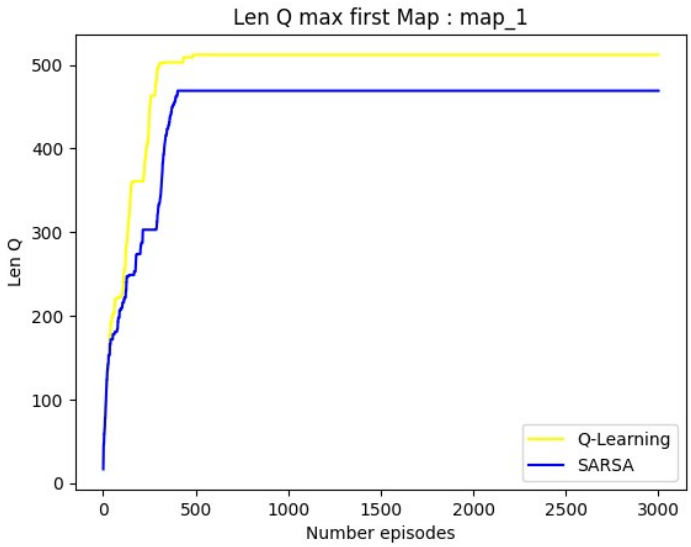


Acest lucru se intampla deoarece o bucata de branza este aproape de soarece, si poate sa o ia pana este prins de prisica.

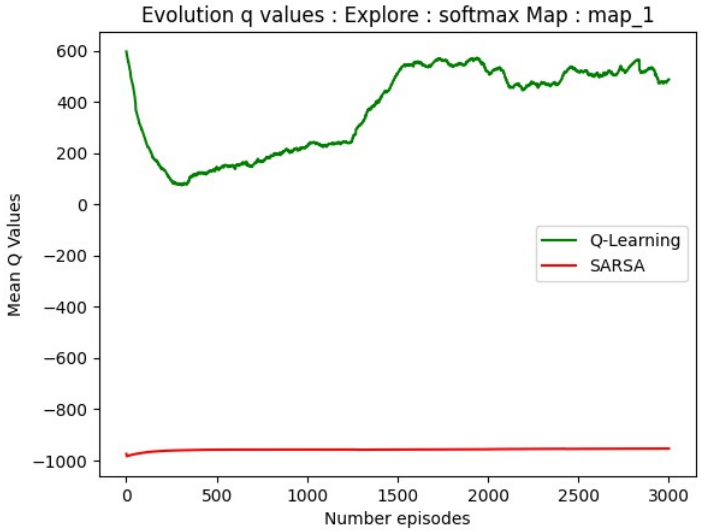
V) Q-Learning vs SARSA:

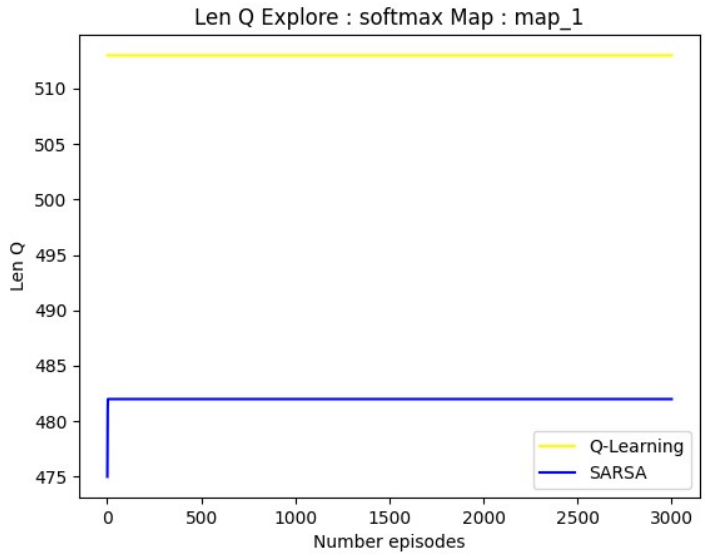
1) Max First:



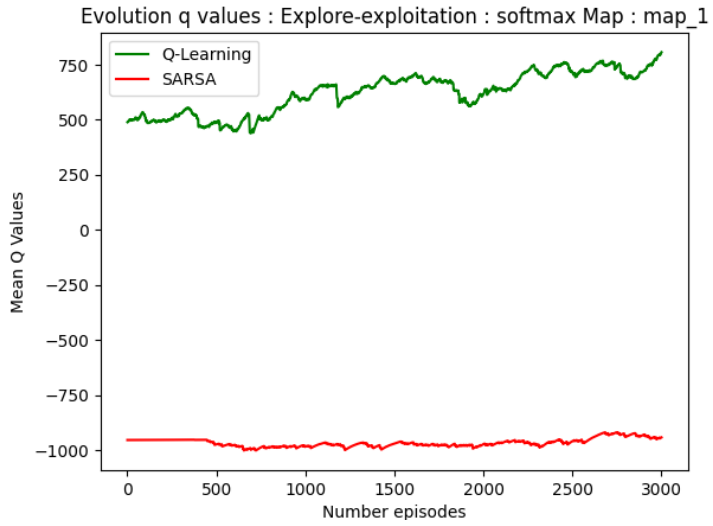


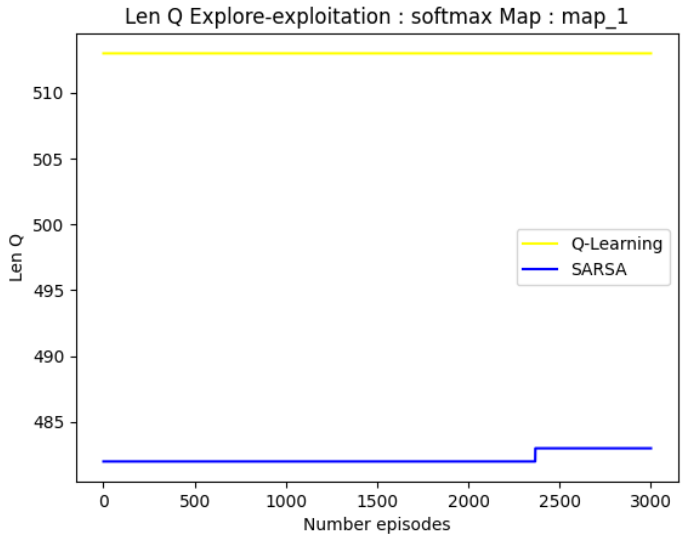
2) Explore: softmax





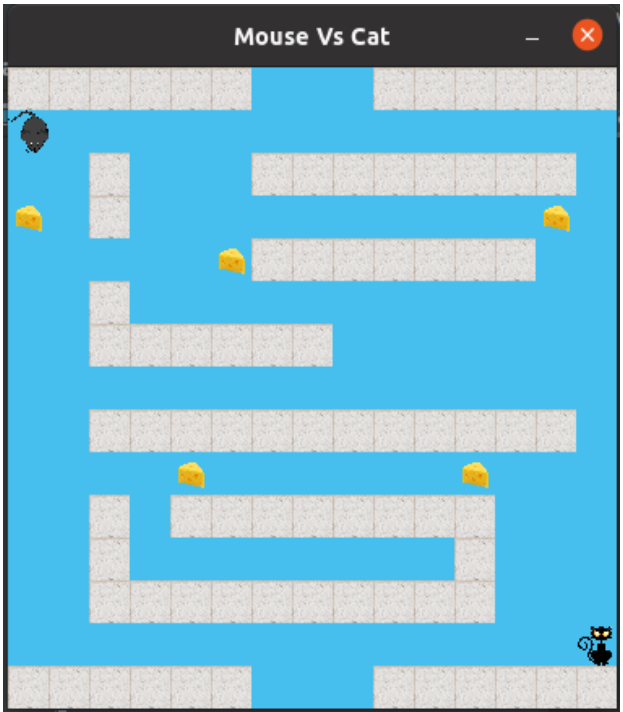
3) Exploatare/Eploare: softmax





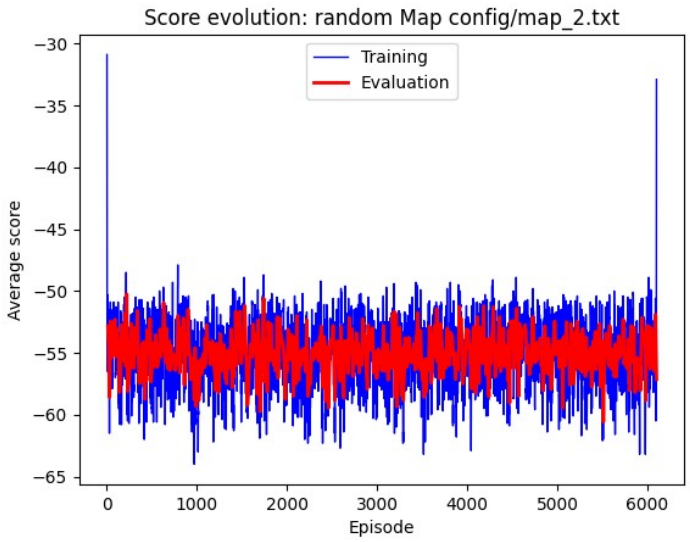
Map 2:

Numar de rulari: 6100

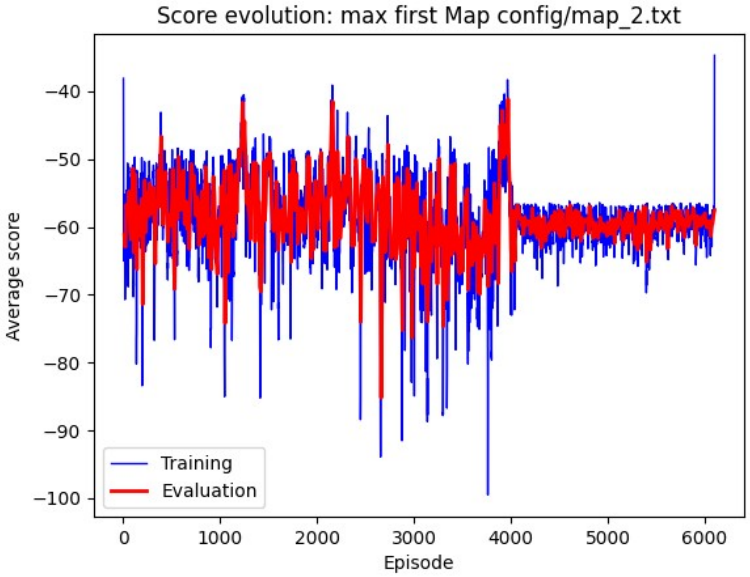


II) Evolutia scorului în funcție de numărul episodului de antrenament:

1) Random:

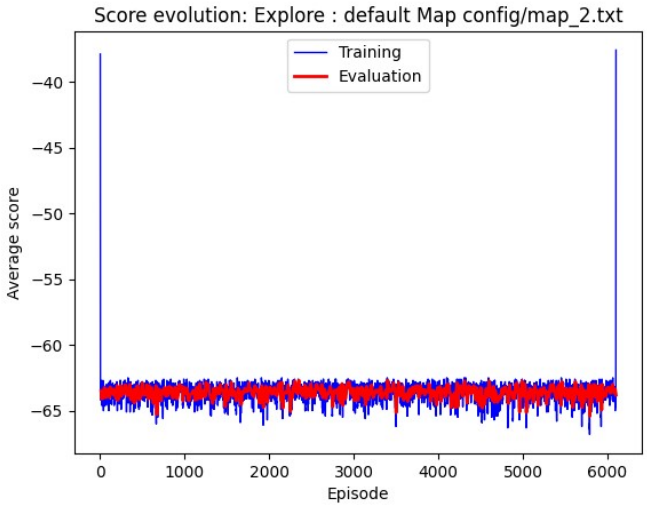


2) Max First:

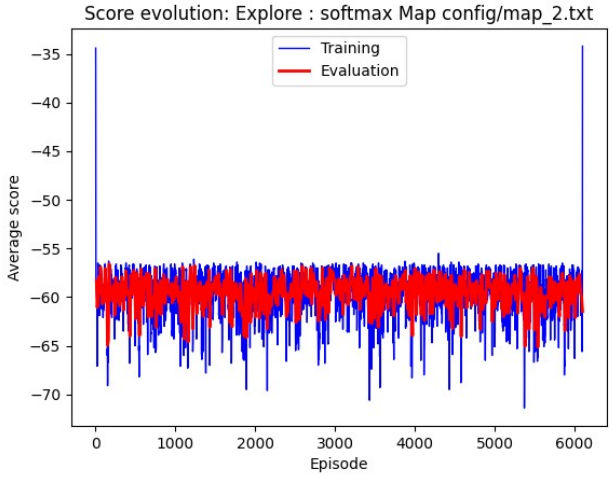


3) Explore:

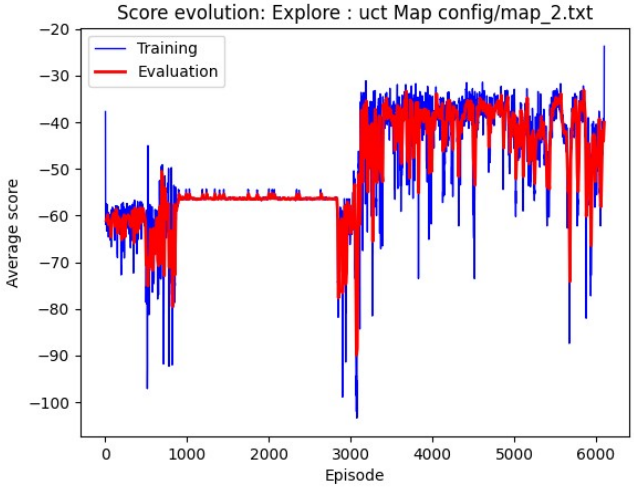
a) Default:



b) Softmax:

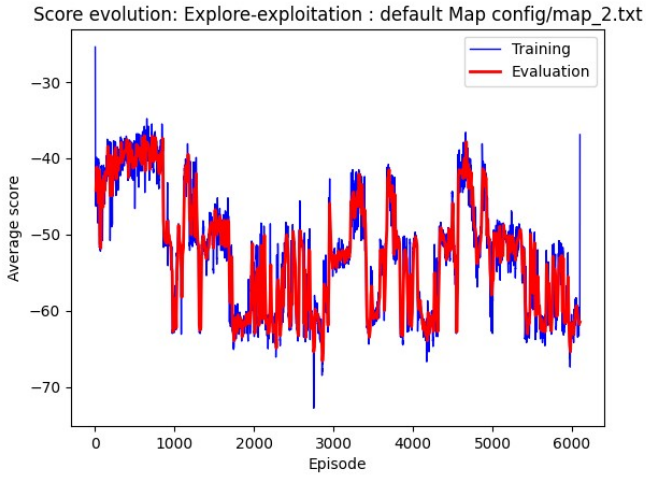


c) UCT:

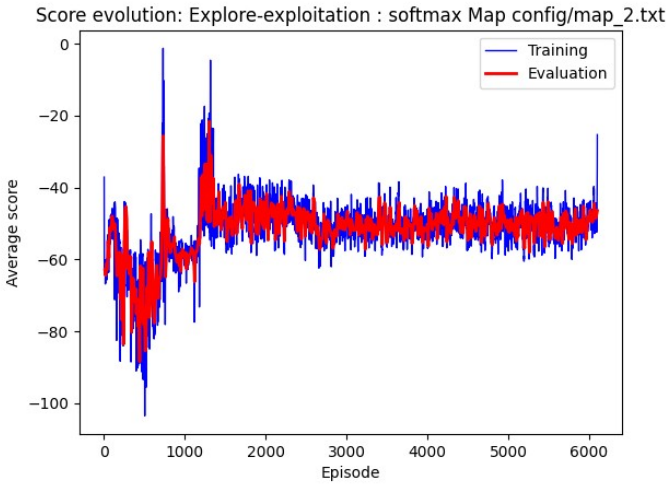


3) Exploatare/Explorare:

a) default:



b) softmax:



II) Evolutia scorului în funcție de numărul episodului de antrenament:

1) Random:

D.F. / L.R	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.9	0	0	0	0	0	0	0
0.85	0	0	0	0	0	0	0
0.8	0	0	0	0	0	0	0
0.75	0	0	0	0	0	0	0
0.7	0	0	0	0	0	0	0
0.65	0	0	0	0	0	0	0
0.5	0	0	0	0	0	0	0

2) Max First:

D.F. / L.R	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.9	0.0008	0.0008	0	0	0.0098	0.2205	0.1193
0.85	0.0013	0.0002	0	0	0	0	0.9202
0.8	0.0013	0.0221	0	0	0.0087	0.12	0.0011
0.75	0.0128	0.0157	0.0003	0.0002	0.002	0.0007	0.0079
0.7	0.0007	0.0034	0	0.0018	0.0025	0	0.0152
0.65	0.0074	0	0.0002	0	0.9666	0.0026	0.0011
0.5	0.0613	0.0003	0.0002	0.0044	0	0	0

3) Explore:

a) Default:

D.F. / L.R	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.9	0	0	0	0	0	0	0
0.85	0	0	0	0	0	0	0
0.8	0	0	0	0	0	0	0
0.75	0	0	0	0	0	0	0
0.7	0	0	0	0	0	0	0
0.65	0	0	0	0	0	0	0
0.5	0	0	0	0	0	0	0

b) SoftMax:

D.F. / L.R	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.9	0.109	0.2062	0.0384	0	0.0769	0.0123	0.1736
0.85	0.1451	0.0402	0.0508	0.0152	0.0054	0.0669	0.0105
0.8	0.143	0.0367	0.0108	0.142	0.0082	0.0095	0.0128
0.75	0.1451	0.0666	0.0026	0.0449	0.0769	0.0207	0.0346
0.7	0.1228	0.03	0.0513	0.0021	0.0315	0.0039	0.0033
0.65	0.0656	0.0346	0.0462	0.0192	0.0416	0.0341	0.0033
0.5	0.1043	0.0331	0.0043	0.0103	0.0256	0.0521	0.0089

4) Eploare/Exploatare:

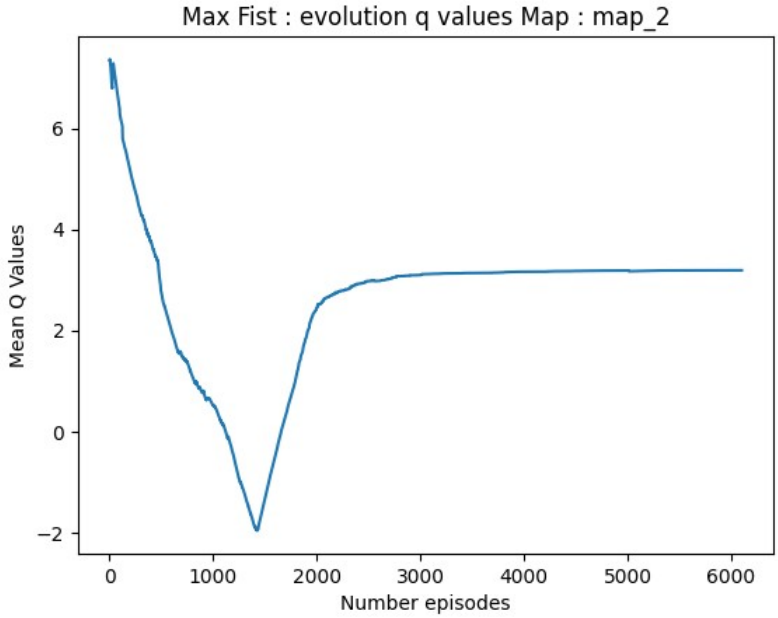
a) default:

D.F. / L.R	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.9	0	0	0.0025	0	0	0	0
0.85	0.2159	0	0	0	0	0	0
0.8	0	0	0	0	0	0	0
0.75	0	0	0	0	0	0	0
0.7	0	0	0	0	0	0	0
0.65	0	0	0	0	0	0	0
0.5	0	0	0	0	0	0	0

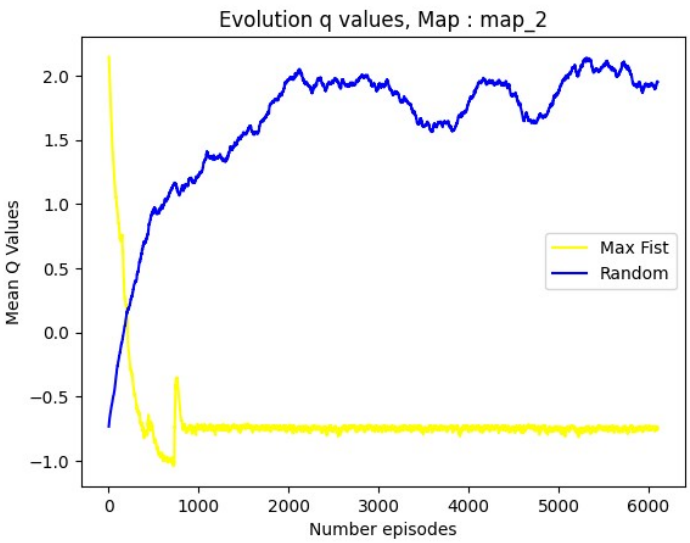
b) softmax:

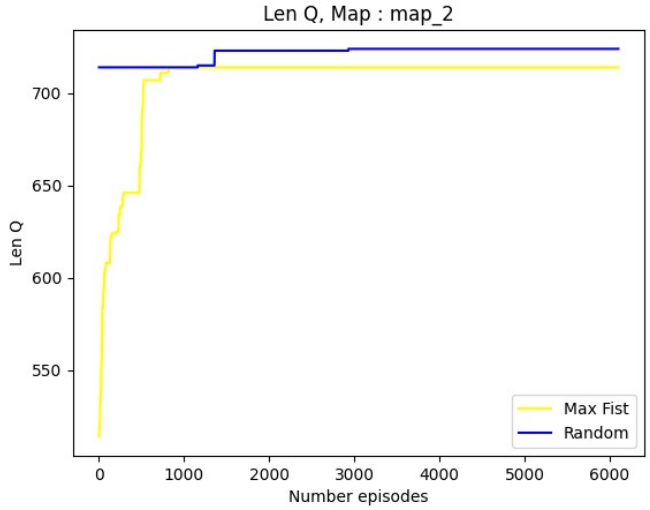
D.F. / L.R	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.9	0.1462	0.0905	0.1567	0.0928	0.041	0.0444	0.0818
0.85	0.1334	0.0518	0.029	0.0259	0.0028	0.0505	0.0254
0.8	0.1003	0.0187	0.0702	0.0195	0.0093	0	0.0139
0.75	0.067	0.0685	0.0044	0.0133	0.0067	0.0262	0.032
0.7	0.1	0.0584	0.0013	0.0166	0.0466	0.0243	0.0111
0.65	0.1326	0.0269	0.0079	0.0421	0.0098	0.0054	0.012
0.5	0.0707	0.0234	0.0103	0.0951	0.0377	0.0207	0.031

III) Cum afectează numărul de episoade de antrenament valorile din tabela de utilitati în cazul strategiei max first?



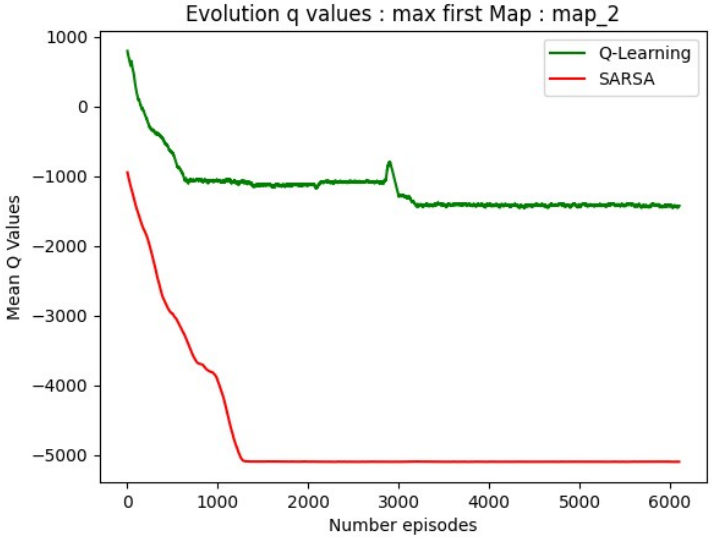
IV)Care sunt diferențele intre tabela de utilitati din cazul strategiei max first și random?

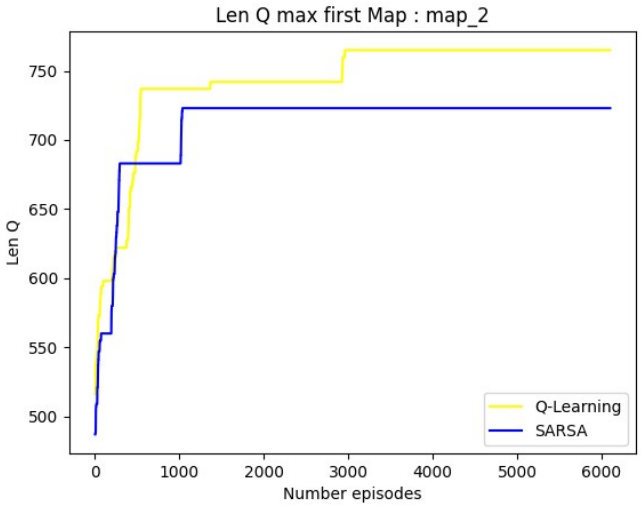




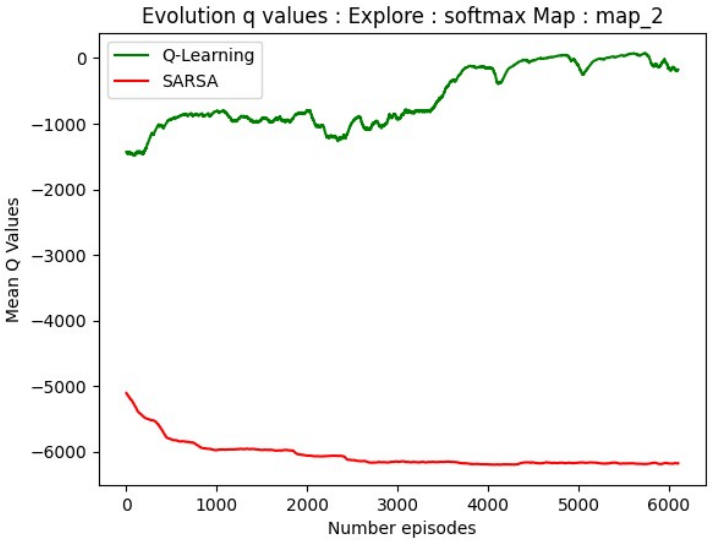
V) Q-Learning vs SARSA:

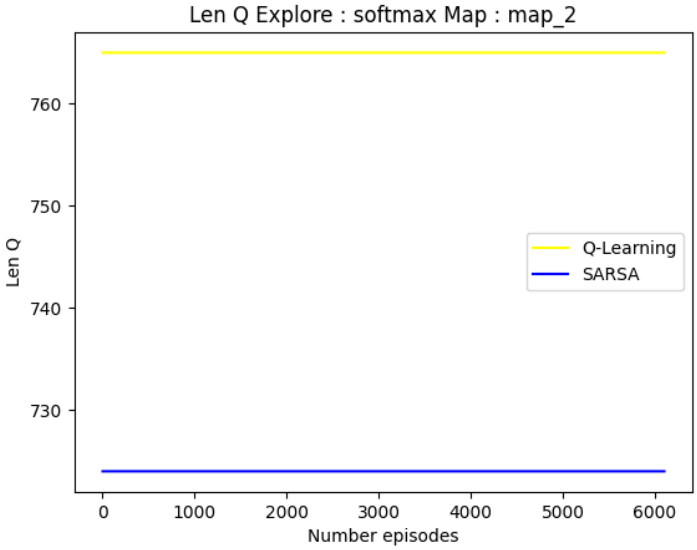
1) Max First



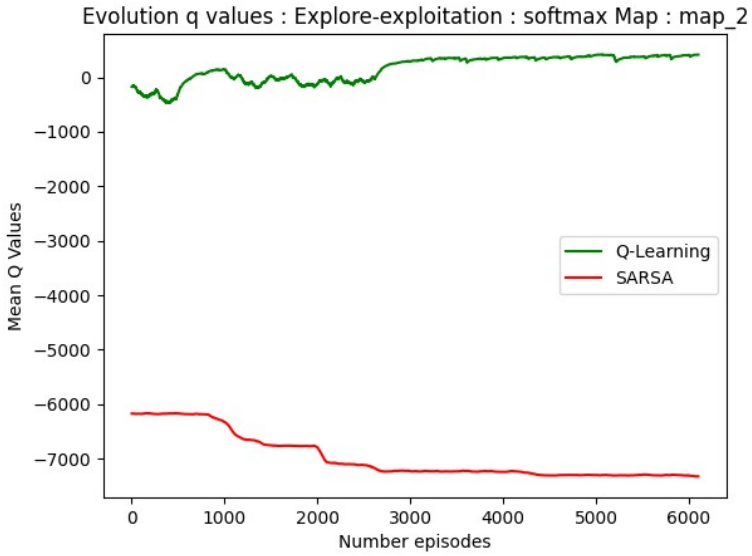


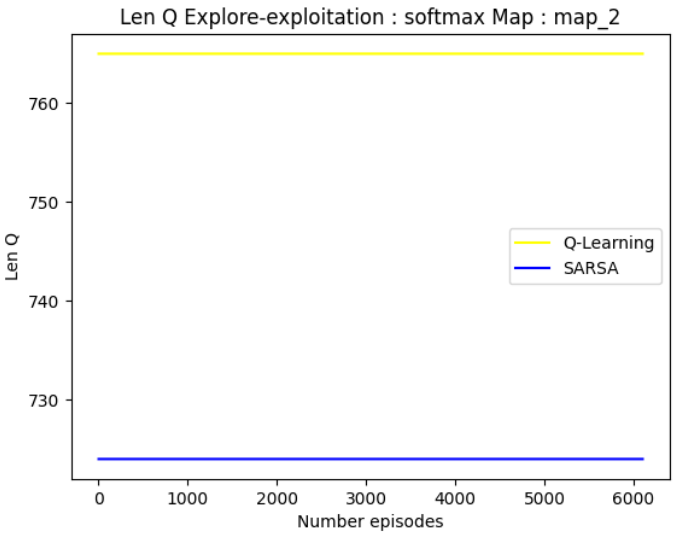
2) Expore: Softmax





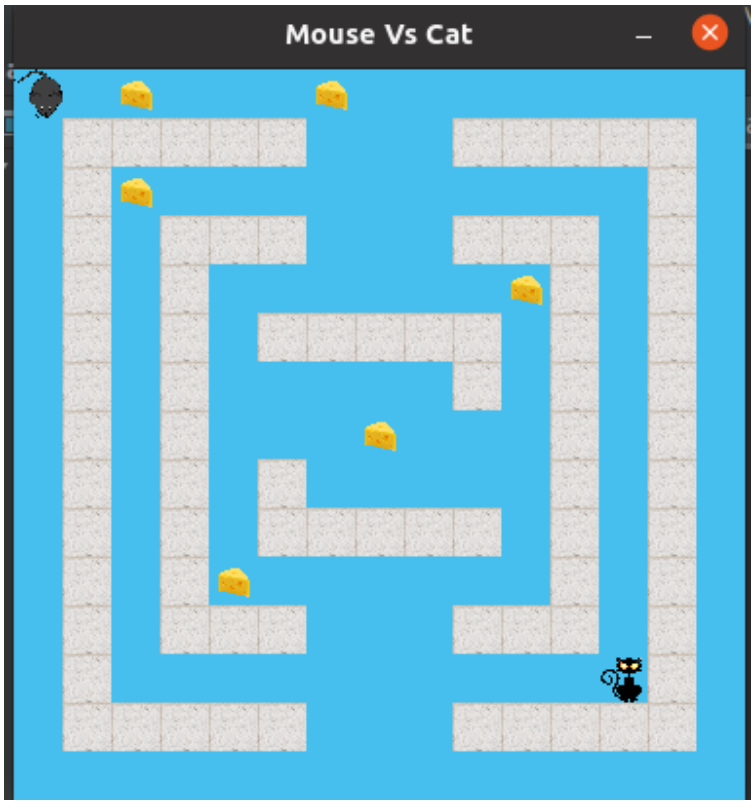
3) Explore/Exploatation: softmax





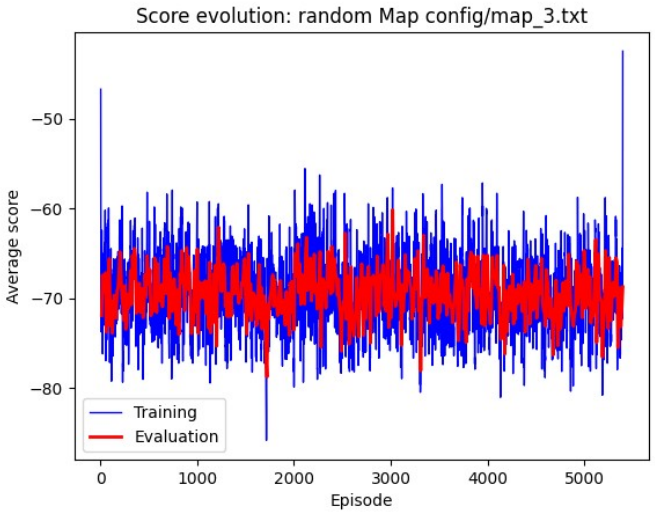
Map3 :

Numar Rulari : 5400

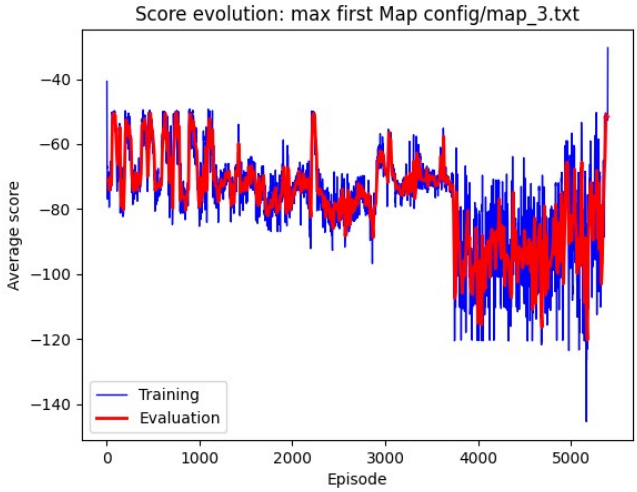


I) Evolutia scorului în funcție de numărul episodului de antrenament:

1) Random:

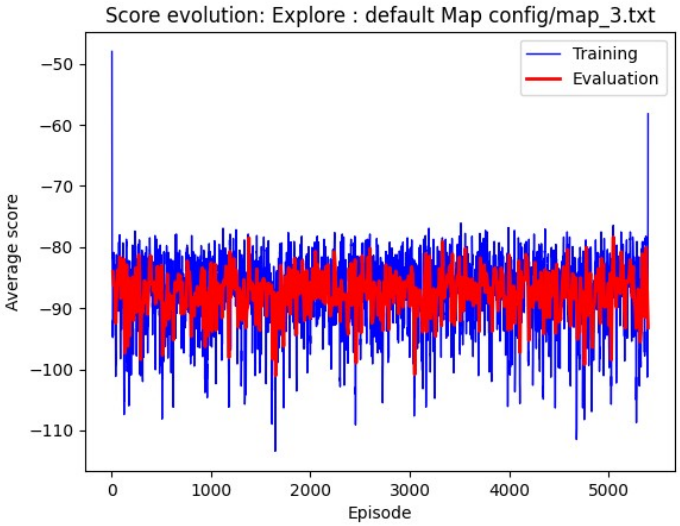


2) Max First:

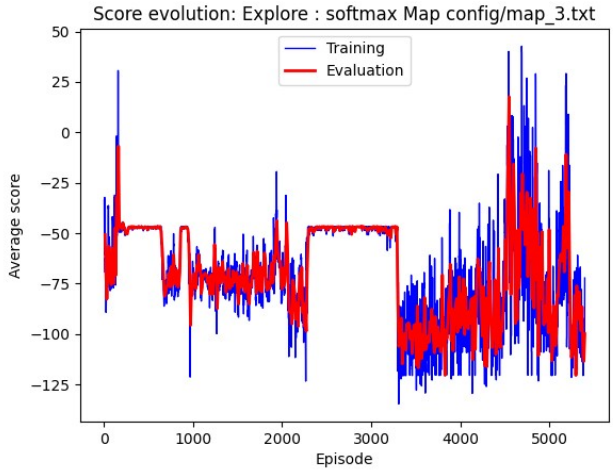


3) Explore:

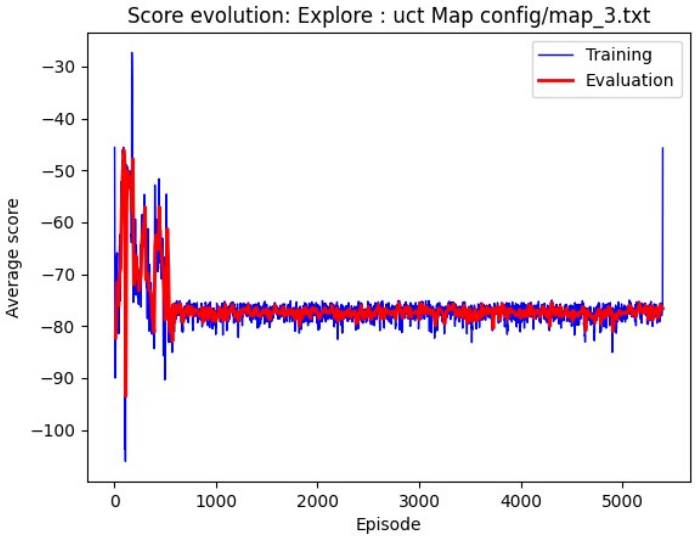
a) default:



b) softmax:

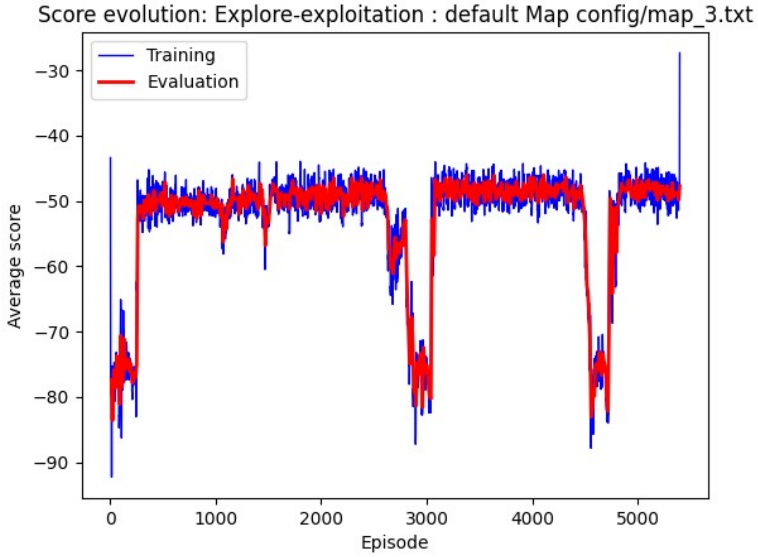


c) uct:

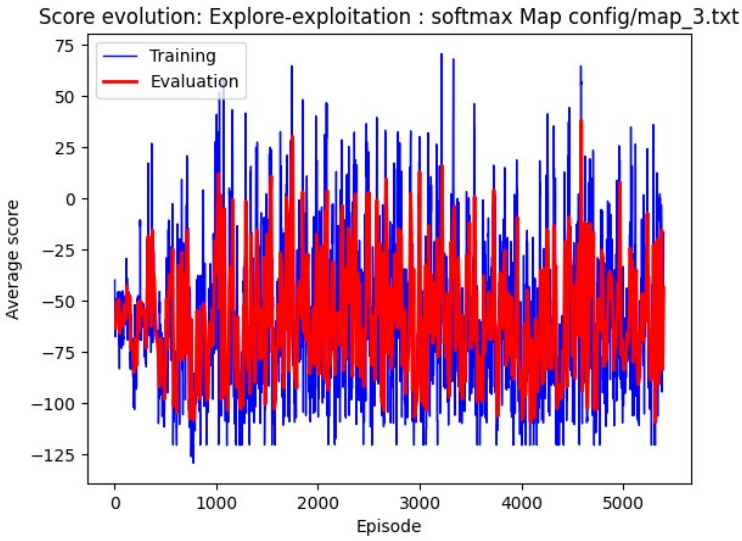


4) Exploatare/Eploreare:

a) defautl:



b) softmax:



II) Evolutia scorului în funcție de numărul episodului de antrenament:

1) Random:

D.F. / L.R	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.9	0	0	0	0	0	0	0
0.85	0	0	0	0	0	0	0
0.8	0	0	0	0	0	0	0
0.75	0	0	0	0	0	0	0
0.7	0	0	0	0	0	0	0
0.65	0	0	0	0	0	0	0
0.5	0	0	0	0	0	0	0

2) Max First:

D.F. / L.R	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.9	0.0015	0.0407	0.073	0.9502	0.0022	0.0259	0.0183
0.85	0.0839	0.4948	0.0015	0.008	0.9622	0.9035	0.0019
0.8	0.3576	0.1181	0.2098	0.8744	0.9144	0.0117	0.0015
0.75	0.8713	0.0598	0.223	0.9481	0.9409	0.9474	0.9557
0.7	0.2672	0.9257	0.9343	0.2044	0.057	0.8937	0.938
0.65	0.2783	0.0828	0.9485	0.972	0.1226	0.8907	0.0476
0.5	0.2104	0.9289	0.2011	0.138	0.8935	0.6783	0.0339

3) Explore:

a) default:

D.F. / L.R	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.9	0	0	0	0	0	0	0
0.85	0	0	0	0	0	0	0
0.8	0	0	0	0	0	0	0
0.75	0	0	0	0	0	0	0
0.7	0	0	0	0	0	0	0
0.65	0	0	0	0	0	0	0
0.5	0	0	0	0	0	0	0

b) softmax:

D.F. / L.R	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.9	0.4517	0.3374	0.3456	0.2274	0.3285	0.2385	0.0763
0.85	0.4941	0.4065	0.3752	0.4157	0.382	0.2081	0.3011
0.8	0.4722	0.4187	0.3931	0.4087	0.3644	0.2815	0.3628
0.75	0.5065	0.5187	0.3763	0.4861	0.4739	0.3381	0.3167
0.7	0.5331	0.4469	0.4439	0.472	0.4237	0.478	0.3248
0.65	0.5411	0.4859	0.4198	0.4611	0.4578	0.4896	0.4046
0.5	0.4454	0.4304	0.4656	0.3781	0.3985	0.3852	0.3656

4) Eploatare/Explorare:

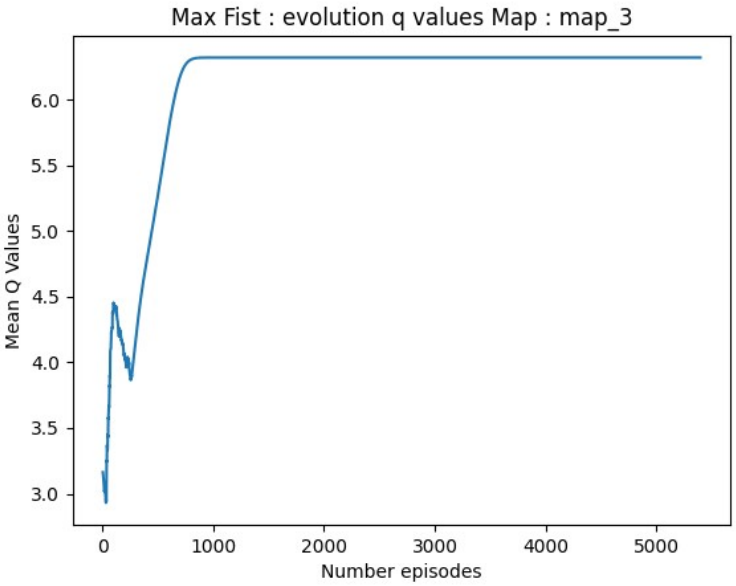
a) default:

D.F. / L.R	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.9	0	0	0	0	0	0	0
0.85	0	0	0	0	0	0	0
0.8	0	0	0	0	0	0	0
0.75	0	0	0	0	0	0	0
0.7	0	0	0	0	0	0	0
0.65	0	0	0	0	0	0	0
0.5	0	0	0	0	0	0	0

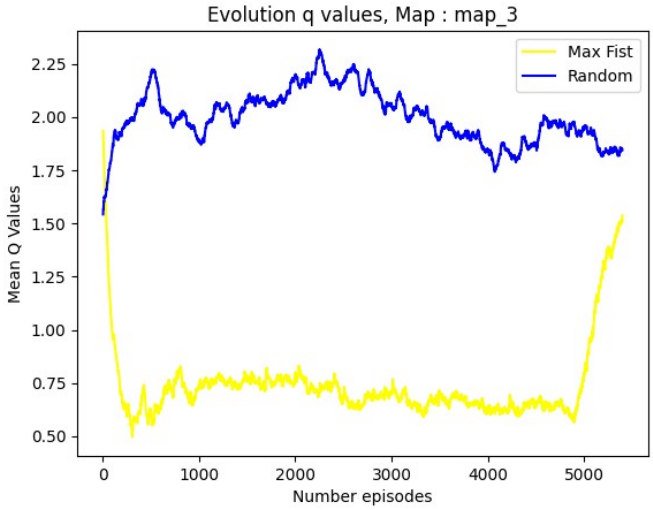
b) softmax:

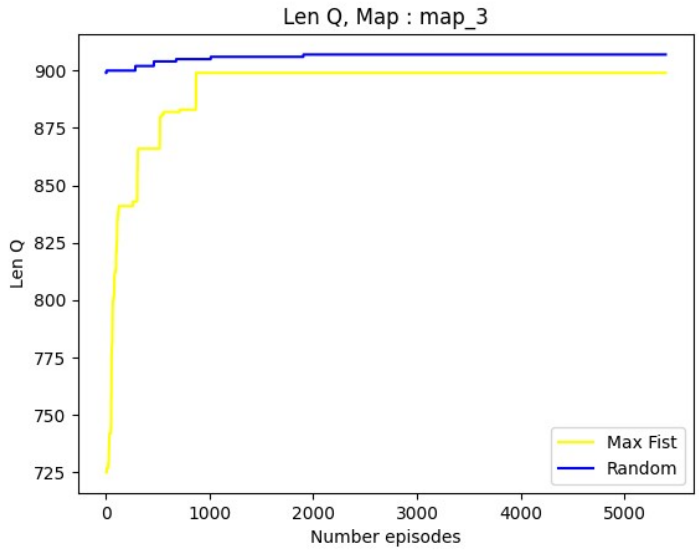
D.F. / L.R	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.9	0.7978	0.3994	0.3963	0.3389	0.0769	0.2631	0.2543
0.85	0.477	0.3824	0.4791	0.1998	0.3759	0.288	0.3344
0.8	0.4398	0.4298	0.3524	0.433	0.4106	0.4231	0.3726
0.75	0.5324	0.3741	0.4409	0.4967	0.4148	0.4498	0.3563
0.7	0.5459	0.5063	0.5111	0.4919	0.4385	0.413	0.465
0.65	0.5356	0.4802	0.5226	0.4585	0.4622	0.4572	0.3826
0.5	0.4309	0.4015	0.4444	0.4044	0.3931	0.4406	0.3933

III) Cum afectează numărul de episoade de antrenament valorile din tabela de utilitati în cazul strategiei max first?



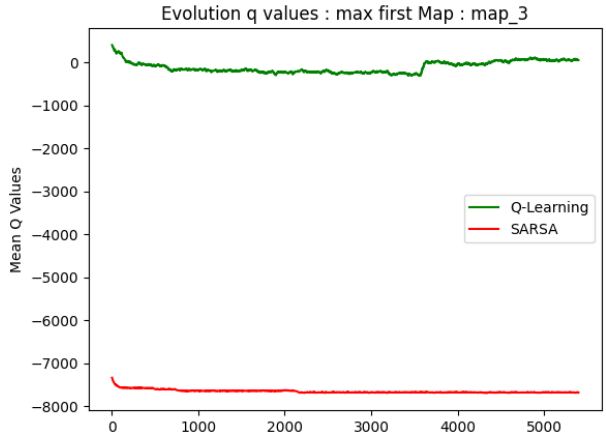
IV)Care sunt diferențele intre tabela de utilitati din cazul strategiei max first și random?



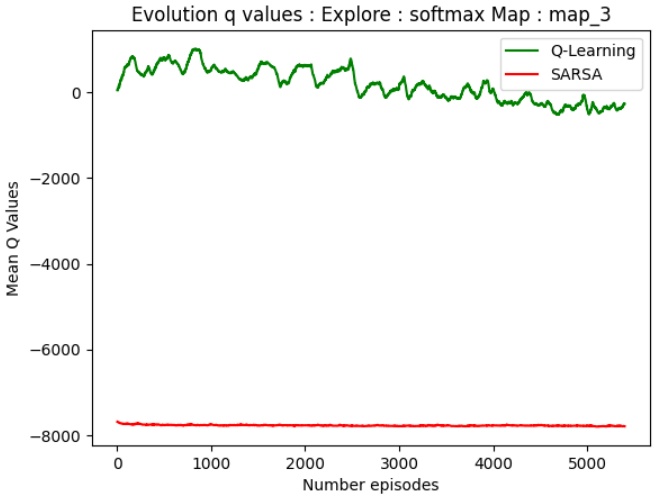


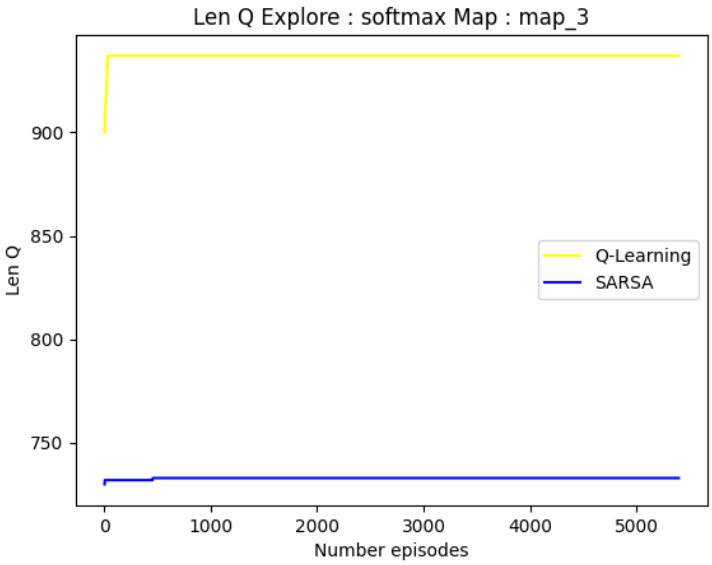
V) Q-Learning vs SARSA:

1) MaxFirst:



2) Explore: softmax





3) Eplorare/Eploatare: softmax: