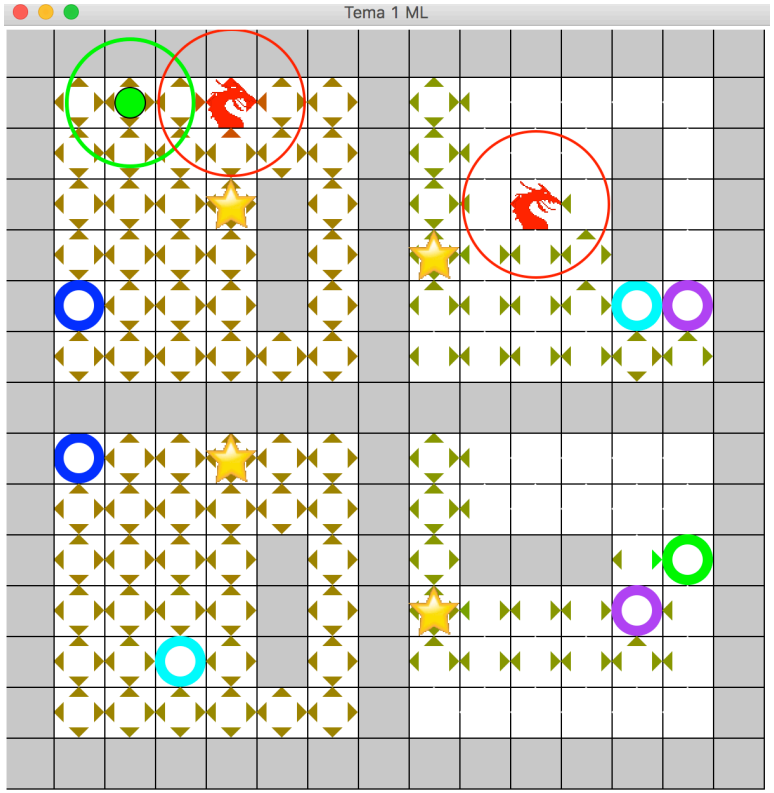
Friday, 7 April 2017

Tema 1 ML Serban Alexandru 341C5

Aceasta tema presupune parcurgerea unui labirint generat aleator de catre eroul nostru Gigel, acesta are ca scop gasirea portalului verde EXIT, pentru a ajunge la acesta, Gigel trebuie sa exploreze harta. Folosind algoritmul QLearning trebue sa il ajutam pe Gigel sa evite gardienii si sa invete cel mai bun drum catre iesire intr-un numar minim de incercari (episoade)

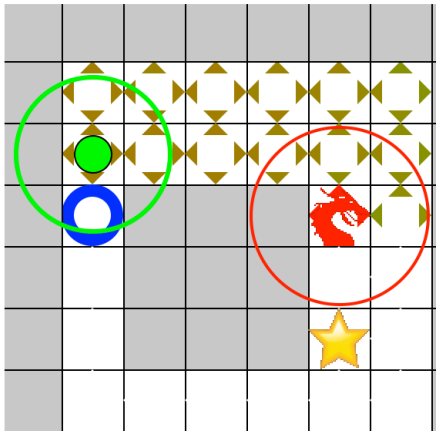


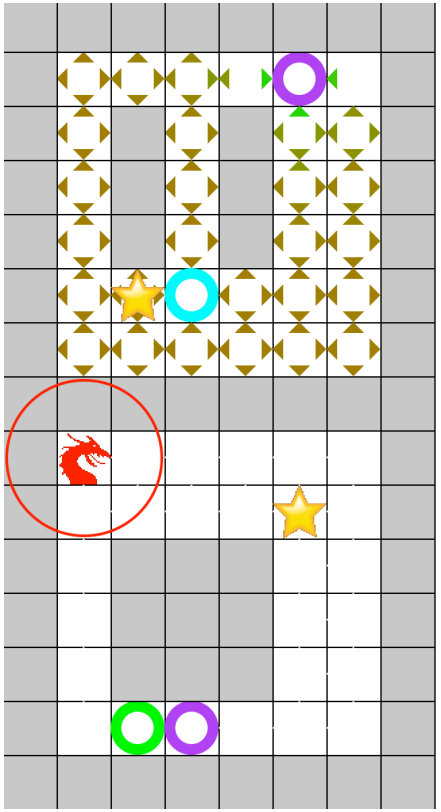
Jocul :

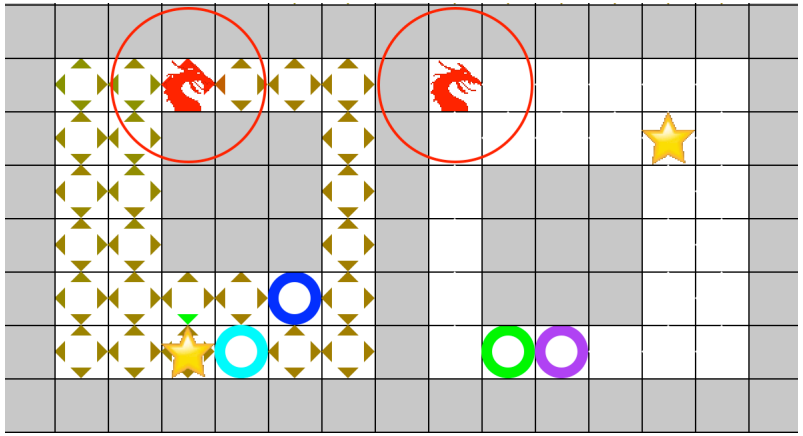
- pentru a reprezenta jocul cat mai subtil am folosit Tkinter
- Algoritmul de QLearning ruleaza intr-un thread sepearat
- Argumentele jocului :
  - Se pot seta viteza jocului (perioada de sleep intre mutari),
  - epsilon pentru strategia jucatorului,
  - numar maxim de episoade
  - strategiile de joc pentru Gigel si Gardieni
  - plot daca dorim sa ne afiseze graficul pentru a vedea cat de repede a invatat Gigel
  - fix_map pentru a face mai multe rulari pe aceasi configuratie de harta

Harta :

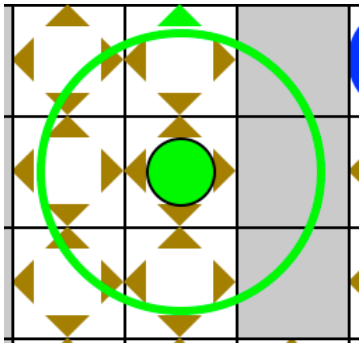
- este generata aleator la pornirea jocului
- atat peretii cat si obiectele de pe harta sunt generate aleator
- este impartita in 4 zone : cadranul/lumi/harti 1,2,3,4 in cadranul 4 se afla EXIT iar in cadranul 1 se afla Gigel
- in fiecare cadran poate exista maxim un gardian
- in fiecare cadran exista minim 1 portal care comunica cu un alt cadran, portalele sunt generate pentru ca gigel sa poata traversa cadrenele la rand catre iesire si oricand sa se poata intoarce
- locatia lui Gigel este generata aleator in primul cadran departe de orice gardian (ca sa nu se blocheze jocul)
- harta este retinuta atat sub forma unei matici cat si sub forma de perechi de coordonate pentru locatia peretilor
- este asigurata generarea unui punct aleator de spawn care nu se suprapune cu un alt obiect de pe harta
- fiecare state(pozitie disponibila) are atribuita 4 actiuni (sus jos stanga dreapta)
- si fiecare actiune are atribuita o utilitatea, reprezentata de un tringhi se semnfica directia de deplasarea catre urmatoarea stare si in functie de culoare utilitatea acestei actiuni (mai verde recomnsa mai mare, mai rosu recompensa negativa)



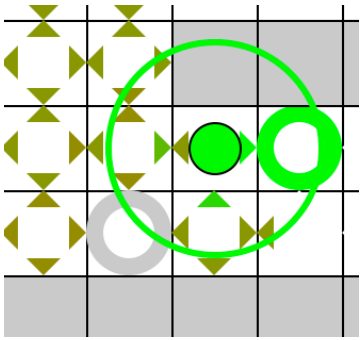




Gigel :



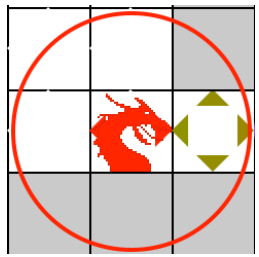
- pozitia lui Gigel este aleasa aleator undeva in cadranul 1 (nu in apropiere de un gardian si intr-un loc disponibil)
- acesta are o raza de actiune ce ii permite sa vadautilitatile actiunilor din jurul lui
- are mai multe strategii :
  - "0 Sta pe loc"
  - "1 Max First” - se depleaseaza in directia cu utilitatea maxima
  - "2 Random” - alege mereu o directie random (NU ESTE INDICAT ca nu invata)
  - "3 Egreedy” - alege rar o actiune random (in functie de epsilon primit) util pentru ca scade sansa de a se bloca dar daca este aleasa o actiune random prea des va ajunge la EXIT dupa un numar mai mare de episoade
  - "4 Manual” - il putem comanda de la tastatura (eventual indruma daca se blocheaza)
  - "5 Smart mark portals” - dezactiveaza portalele prin care trece pentru a micsora spatiul de cautare

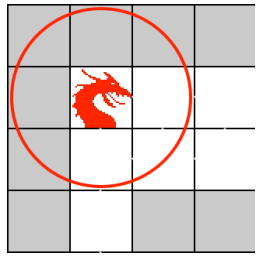


- raza lui de actiune este stabilita in functie de dimensiunea hartii la initializarea jocului
- scorul lui gigel scade la fieare deplasare si la intalnirea cu un gardian
- la gasirea protalului de iesire si a unei recompense scorul sau creste

- mtricea de utilitati este actualizata la fiecare miscare a lui Gigel

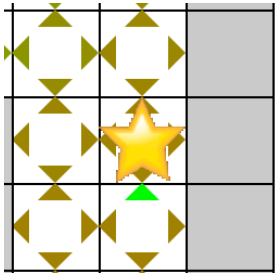
Gardienii : (TODO poza gardian



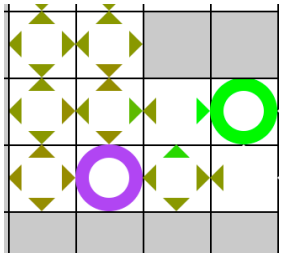


- in fiecare cadran este pus sau nu un gardian cu probabilitate de 1/2 (pot fi 4,3,2,1 sau chiar 0 gardieni)
- fiecare gardian are o raza de vedere care este calculata in fucntei de dimensiunea hartii
- fiecare gardian are o strategie care este setata prin parametrii : strategii :
  - 0 sta pe loc
  - 1 Random (neindicat pentru ca Gigel nu va putea sa invete)
  - 2 se misca pe casuta ce ii confera o distanta mai mica spre Gigel
  - 3 On sight se va misca atunci cand Gigel intra in raza lui de actiune si are o traiectorie libera catre acesta
- gardienii au dreptul sa foloseasca doar spatiile libere de pe harta (nu portale, nu recompense, nu zid)
- gardianul muta primul (apoi Gigel) in functie de strategia aleasa
- daca mutarea unui gardian se suprapune cu pozitia lui Gigel jocul este restartat
- un gardian muta daca Gigel este in raza lui de actiune si in acelasi cadran cu el
- cand un gardian muta deasupra lui Gigel acesta ii da un reward/recompensa negativ(a)

Recompense/Stars :



- in fiecare cadran este generata aleator o recompensa pe care Gigel o poate lua in drumul sau spre iesire pentru a isi maximiza profitul (daca ocolul pentru a o lua nu impune un drum mai costitsitor decat valoarea recompensei)
- o stea/recompensa poate fi colectata o singura data intr-un epsiod, dupa ce Gigel colecteaza o stea acesta o dezactiveaza
- recompensele sunt dezactivate pentru a nu crea o bucla de invatare in care scorul creste la infinit (Gigel va reveni in locul in care a fost recompensat)
- pentru reprezentarea recompenselor am folosit imaginea unei stelute (am incarcat imaginea cu ajutorul interfetei grafice oferite de Tkinter)



Etapa de restart :

- in etapa de restart fiecare obiect ete redesenat in locatia initiala (player devili recompense)
- recompensele sunt reactivate
- portalele sunt reactivate
- resetarea scorului lui Gigel la 1

Analiza Performantelor si Grafice

Pentru analiza performantei algoritmului implementat am folosit un joc de 15x15 cu **aceasi** configuratei de obiecte pentru toate rularile

Pentru a obtine o configuratie fixa se seteaza —map_fix = 1 la parametrii astfel se va incarca harta din fisierul map.txt care a fost creat cu o rulare inainte (este necesar sa facem macar o rulare cu —map_fix = 0 pentru a se genera o harta)

Vom analiza rata de invatare pentru urmatoarele 8 cazuri :

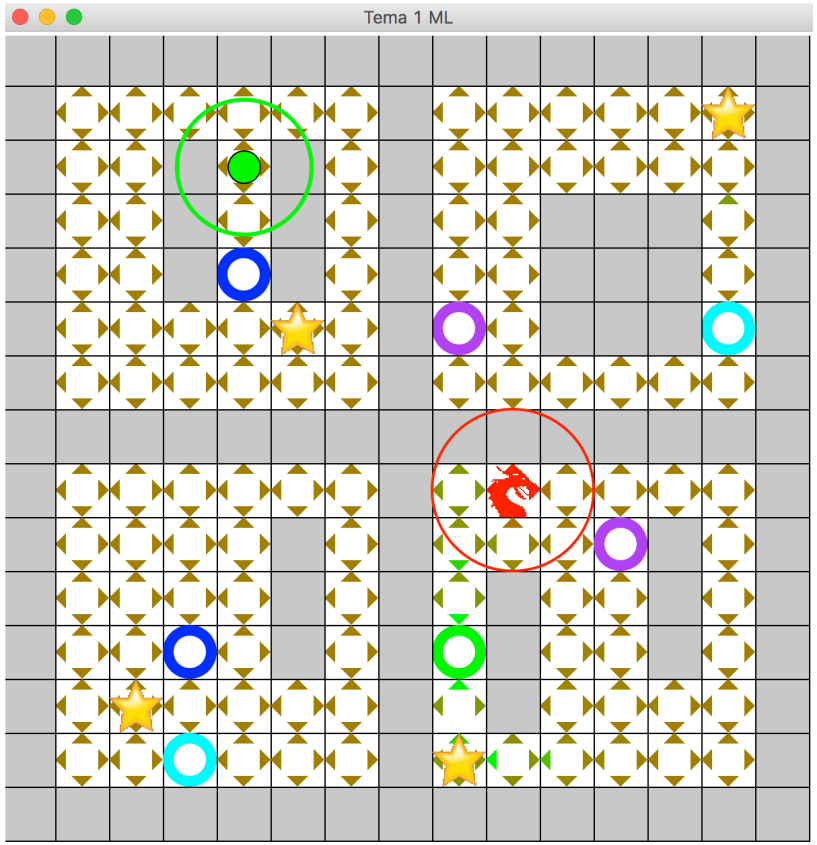
MaxFirst

Random

Egreedy

Smart Mark Portal

si pentu fiecare strategie vom analiza cat de repede invata Gigel atunci cand gardienii se misca respectiv cand ei stau pe loc



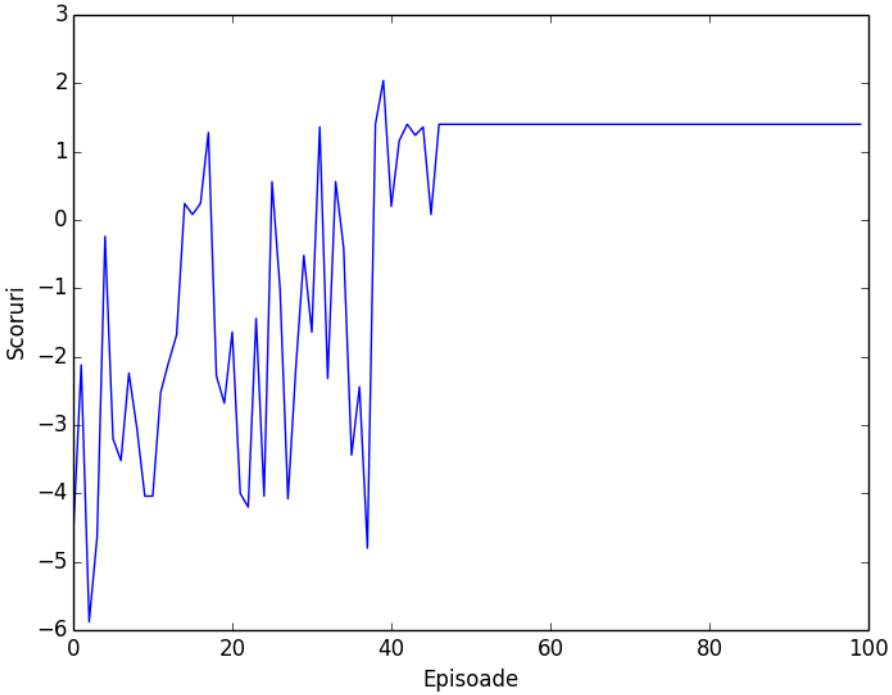
Configuratia fixa pe care am facut toate testele este aceasta

iar parametrii pentru urmatoarele 8 cazuri au fost discount 0.3

sleep 0.0001, learning rate 0.1

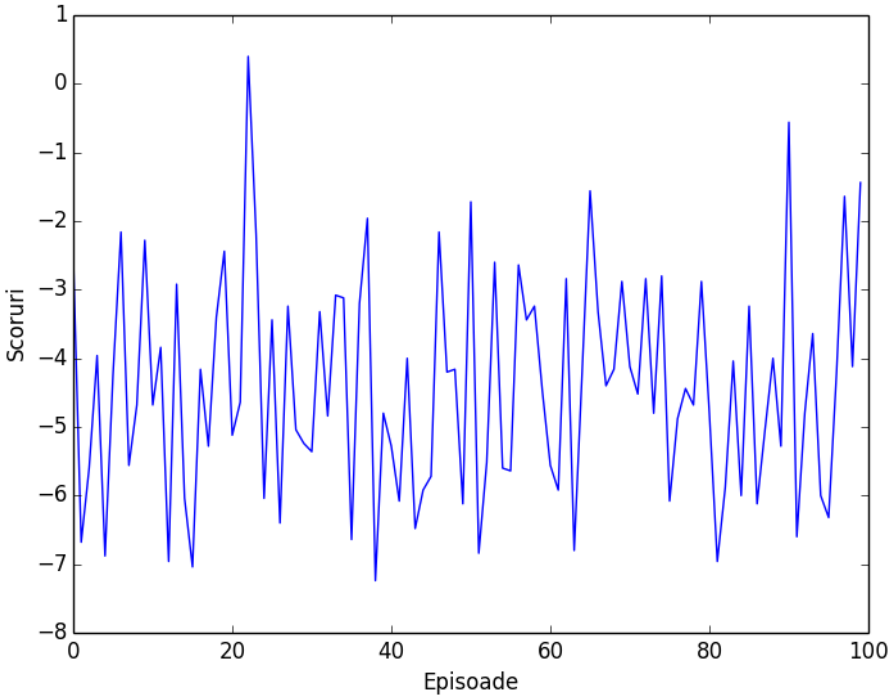
Cazul 1 MaxFirst si Gardieni care stau

In acest caz Gigel alege mereu actiunea cu utilitatea cea mai mare si nu incearca drumuri noi, aceasta varianta este mai stabila dar din cate se poate observa nu ofera scorul maxim, o data gasit un drum desi nu este cel mai optim, este mentinut (poate sa omita recompensele)



Cazul 2 Random si Gardieni care stau SAU Gardieni care se misca

In cazul random Gigel desi se misca in labirint si matricea de utilitati este completata, acesta nu foloseste nimic din cea ce invata pentru ca la fiecare mutare nu tine cont de “ce a invatat”



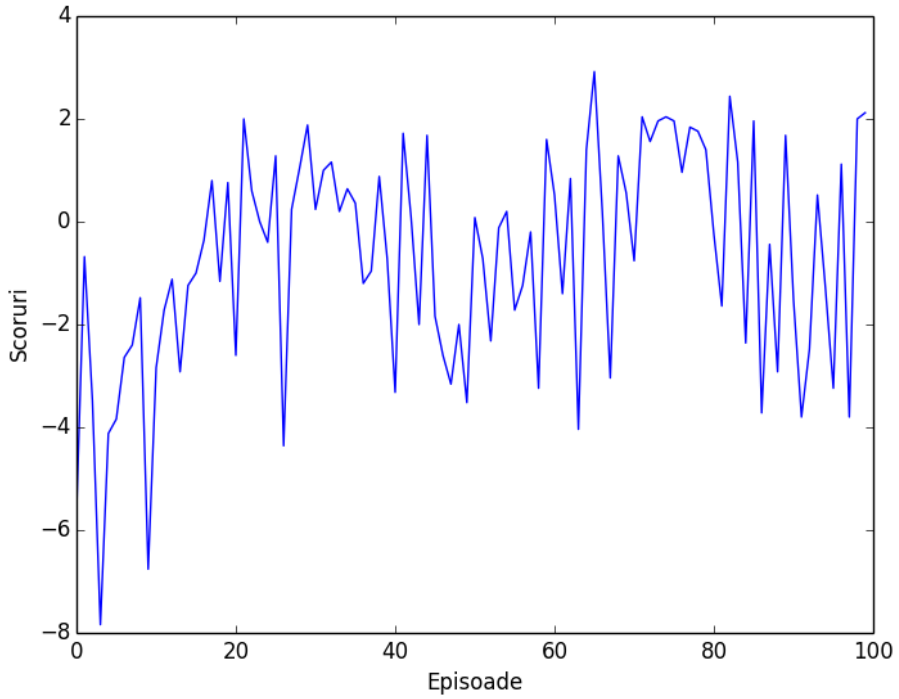
Cazul 3 Egreedy (epsilon 0.12) si Gardieni care stau (1)

In acest caz Gigel invata mai greu deoarece exista o probabilitate 12/100 in care el face o alegere random netinand cont de tabela de utilitati.

Avantajul acestei tehnici este in momentul in care Gigel se poate bloca (spre exemplu atunci cand se intalneste cu un gardian, s-a lovit de el si drumul prin acea casuta il considera foarte neaoptim (utilitate negativa).

D**atorita** miscarii random se duce acolo sidescopera ca gardianul nu mai este in acel loc, schimba utilitatea acelei pozitii si descopera un nou drum, iesind astfel dintr-un eventual impas.

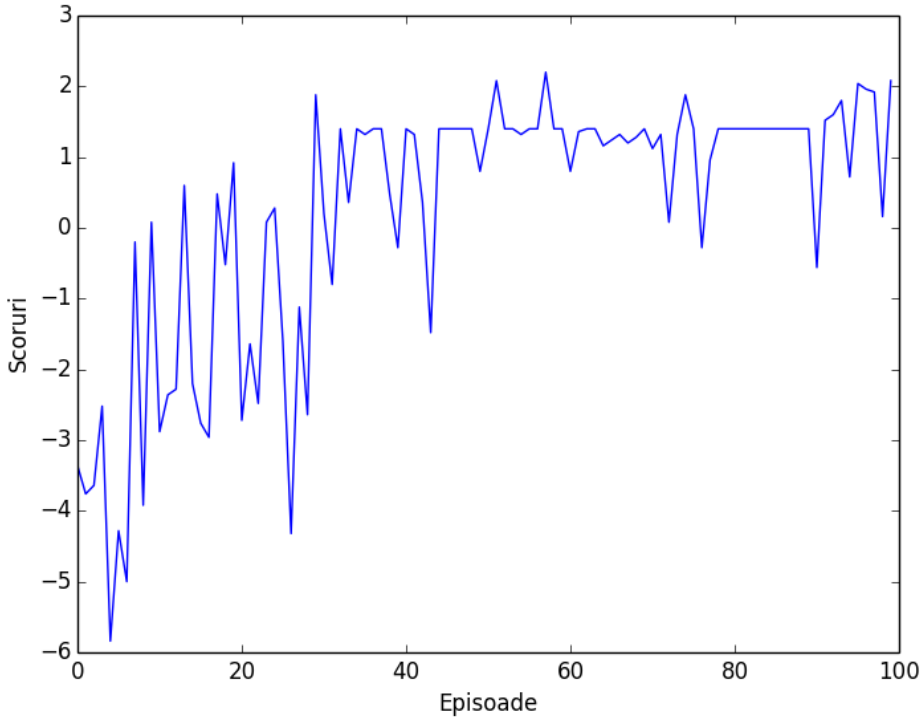
Se poate observa ca pentru un epsilon de 0.12 rata de invatare necesita aproximativ 100 de episoade



In cazul unui epsilon de 0.01 rata de invatare este mai buna, Gigel ajungand sa obtina scoruri pozitive (stabil) dupa episodul 50 (2)

Cu cat rata de alegerea a unei miscari random se micsoreaza algoritmul va deveni mai stabil dar de asemenea poate dura mai mult.

O miscare random poate fi in directia catre un optim global desi este contrara unei miscari MaxFirst



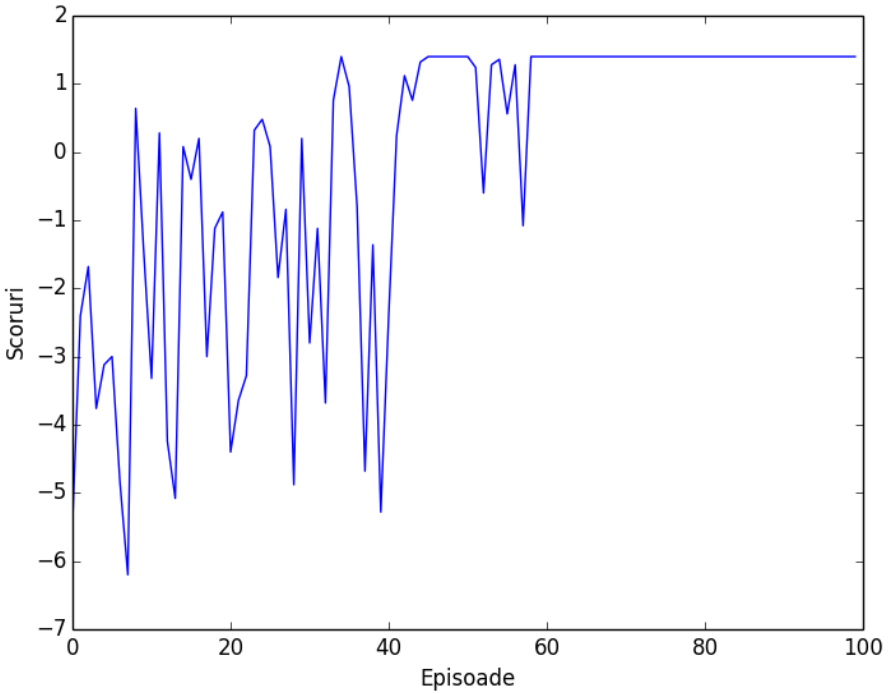
Cazul 4 Smart Portal Mark si Gardieni care stau

In cazul acestei strategii Gigel tine minte portalele prin care acesta a trecut si nu se intoarce, mergand tot inainte catre portalul final acesta reuseste sa invete drumul mai repede datorita numarului mai mic de posbilitati de explorare.

Se poate observa ca in comparatie cu cazul precedent scorurile de final raman constante si se ajunge la un maxim care nu va mai fi depasit, care este pastrat pana la final, desi exista posibilitatea unui trase ce ofera un scor mai bun (se pot omite recompense) O data un drum gasit este pastrat, nu se incearca variante noi.

Pe aceasta configuratie in mod deosebit nu este cu mult mai optim decat Maxfirst dar incercand sa rulez de mai multe ori (pe mai multe harti)

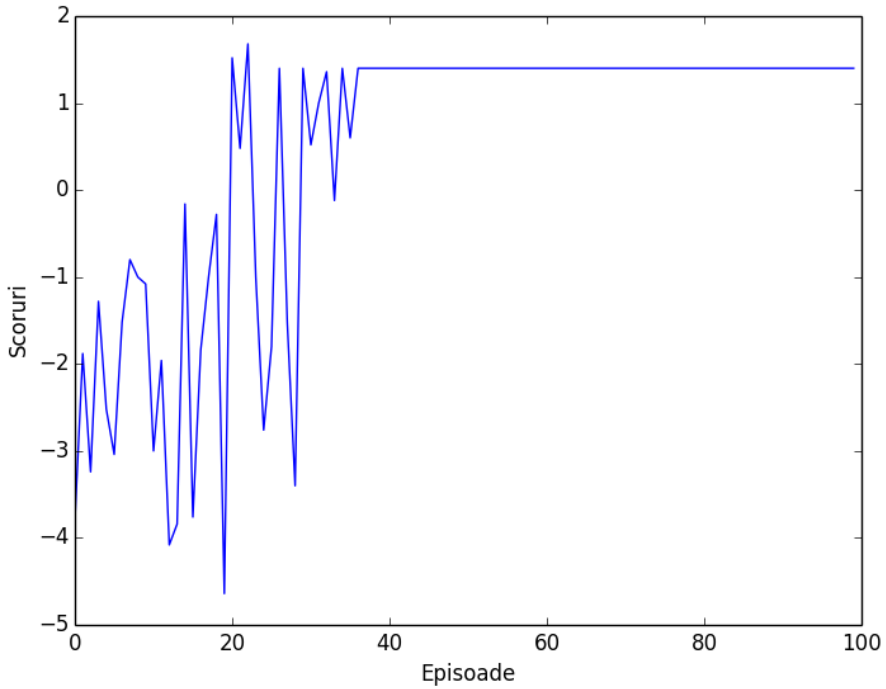
Am putut observa ca a fost de pana la 1.5 ori mai optim.



Cazul 5 MaxFirst si Gardieni care se misca spre Gigel

Eu consider ca in acest caz Gigel s-a intalnit cu gardianul la inceputul jocului si apoi a mers in directia opusa gasind astfel un drum aproximativ la fel de rapid catre portalul de iesire ca si in cazul 1

In acest caz se poate observa ca stabilirea la un scor final fix s-a realizat devreme (inainte de episodul 40) atat datorita algoritumul ce a dictat urmatoarea actiune cu utilitatea maxima cat si datorita amplsamentului obiectelor pe harta



Cazul 6 Random si Gardieni care se misca spre Gigel

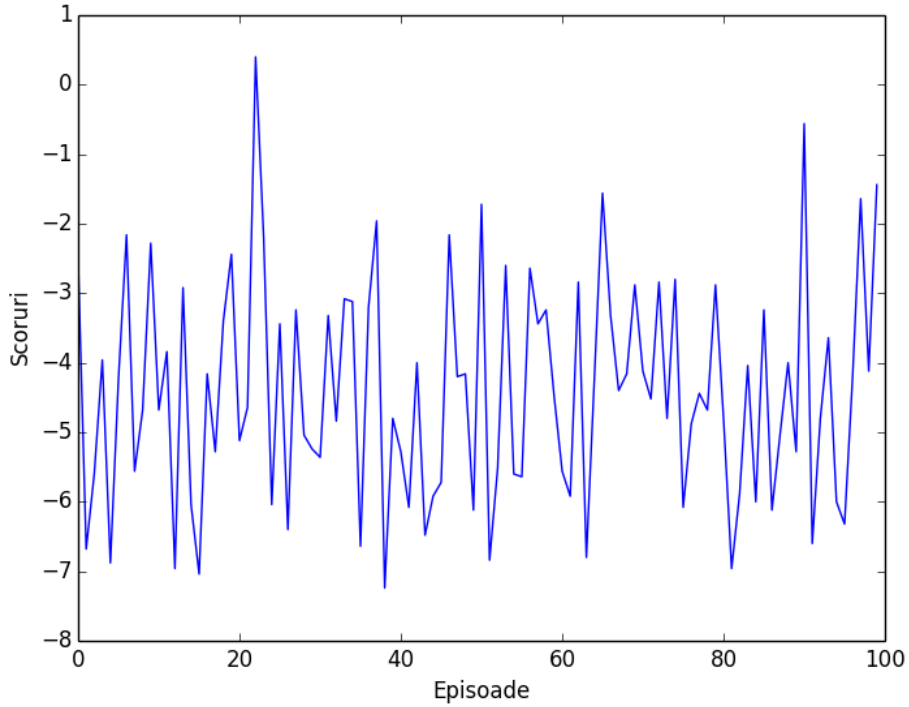
In acest caz Gigel a efectuat fiecare miscare random fara sa tina caont de miscarea cu utilitate maxima dictata de algoritm

Se obtine un rezultat asemanator cazului 2 din care putem observa

- gigel nu a ajuns la un statdiu stabil in care a nvatat drumul

- exista un singur caz in care a fost obtinut un scor pozitiv

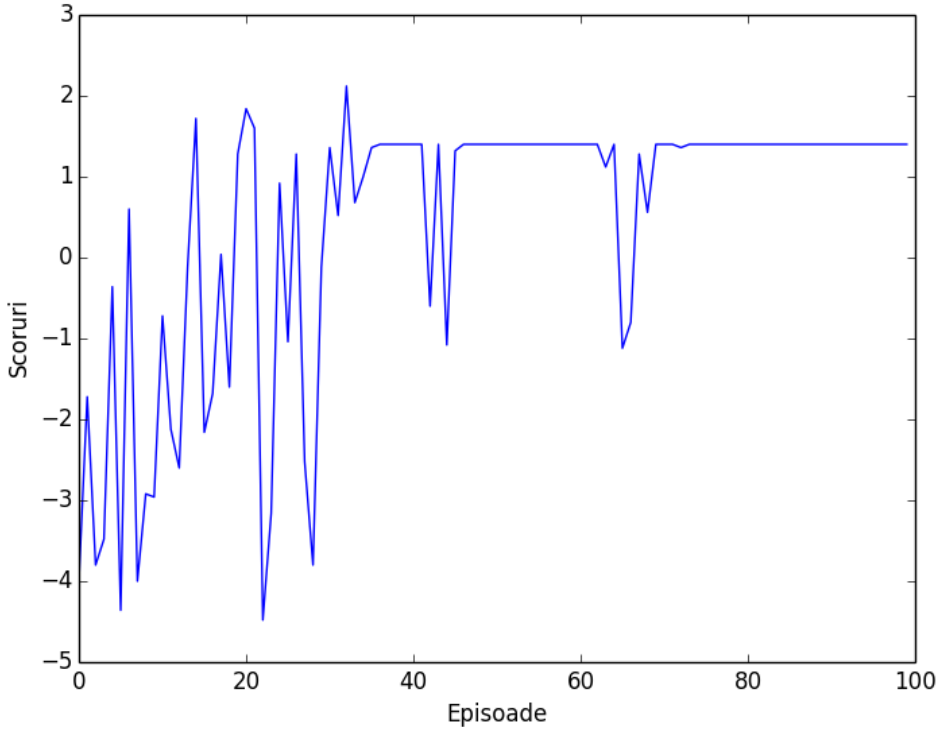
- nu este o modalitate de explorare eficienta



Cazul 7 Egreedy si Gardieni care se misca spre Gigel

In acest caz am folosit epsilon 0.01 iar pozitia gardianului ca si in cazul 6 tind sa cred ca influentat pozitiv evolutia jocului, Gigel invatat drumul si obtinand unscor constant dupa episodul 50.

Desi acesta a gasit drumul spre iesire si l-a urmat aproximativ 15 episoade se poate observat dupa episodul 60 din cauza epsilon care a sugerat o mutare random acesta a avut o scadere in scor

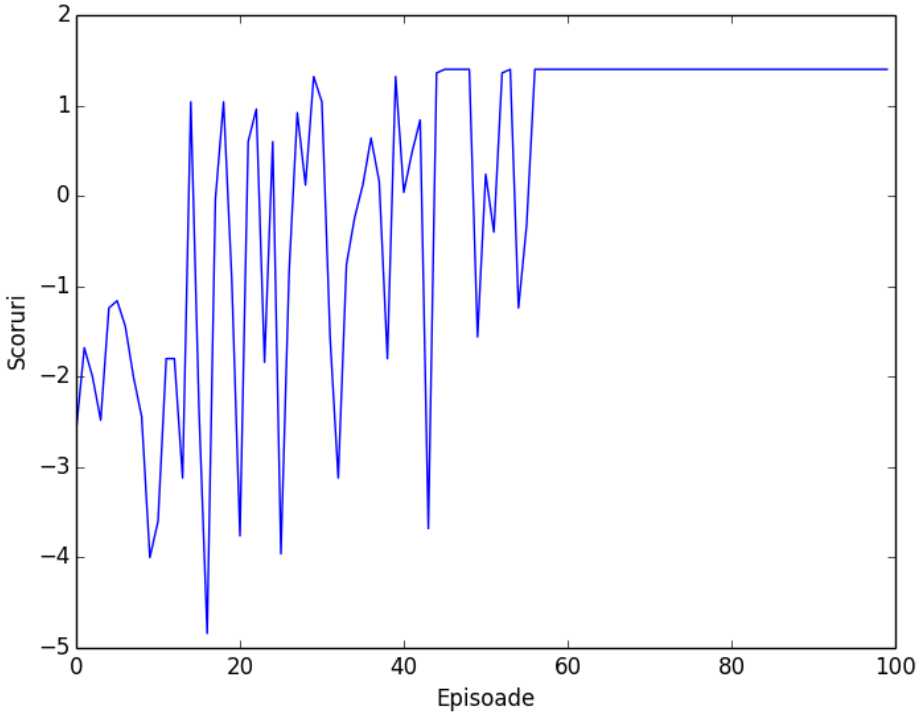


Cazul 8 Smart Portal Mark si Gardieni care se misca spre Gigel

Pozitionarea gardianului in acest caz a incurcat evolutia jocului din cate se poate observa Gigel a invatat drumul optim abia dupa episodul 60.

In urma gasiri drumului optim acesta ramane neschimbat, nu se fac abataeri ca in cazul tehnicii Egreedy.

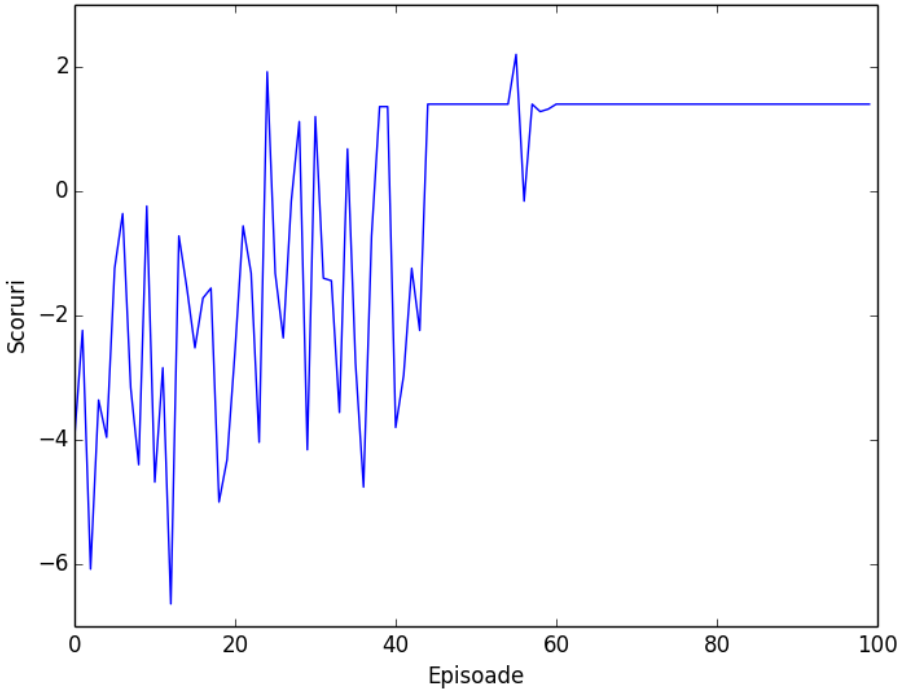
Daca se folosesc tehnici aditionale de oprire din invatare se poate obtine un rezultat mai bun cu tehnica Egreedy decat cu aceasta



Analizare Cazuri in care am schimbat discount si learning rate

Cazul 9 MAXFirst si Gardieni care stau - discount 0.1

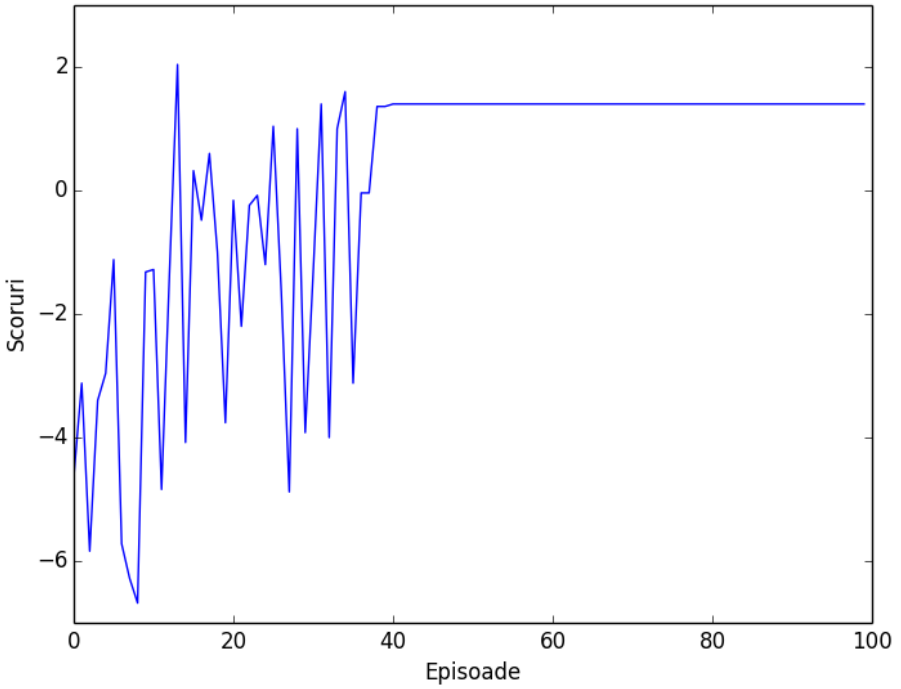
Comparativ cu cazul 1 pot spune ca aceasta este o rata mai rapida de invatare pe aceasta harta in urma folosiri discount 0.1



Cazul 10 MAXFirst si Gardieni care stau - learning rate 0.05 si discount 0.1

In urma schimbarii rate de invatare de la 0.1 la 0.2 am observat in acest context (pe aceasta harta si aceasta configuratei de obiecte) o inbunatatire fata de cazul 9

Din cate se poate observa evolutia scorurilor este asemanatoare dar gigel a invatat (a ajuns la EXIT cu un scor stabil pozitiv) in mai putine epsioade (cu aproximativ 10 episoade mai putin)



Observatii :

Scorul se stabilizeaza in aproximativ 100 de episoade (in cel mai rau caz)

Jocul se poate bloca in cazul in care unul din gardieni alege sa se deplasez in sus si player in jos apoi invers (se poate crea un deadlock) acesta se intampla de obicei atunci cand jucatorul are de ales doar dintre 2 poziti si incearca sa se deplaseze opus fata de gardian (avand utilitate mai mare) iar gardianul incearca sa il prinda

Numarul de camere in cazul abordarii mele este fix (4) dar marimea acestora poate varia crescand exponential complexitatea, spatiul de cautare este mult mai mare iar daca gardienii sunt setati sa urmareasca jucatorul si au o raza mare de actiune vor exista foarte multe situatii in care jucatorul pierde si utilitatea multor actiuni va fi diminuata

Desi numarul de episoade nu este foarte mare un episod poate dura destul de mult pentru ca numarul de operatii este mare si din aceasta cauza am fost limitat la un joc de maxim 15x15

BONUS :

Pe langa interfata vizuala realizata in Tkinter am decis sa fac si partea a2a cea cu portalul, ca indicator am folosit dezactivarea portalului dupa folosirea acestuia

Bibligrafie :

https://www.youtube.com/watch?v=1XRahNzA5bE

laboratorul 5 si materialele, cursul de qlearning

https://en.wikipedia.org/wiki/Learning

ca schelet de cod am pornit de la aceasta implementare https://www.youtube.com/watch?v=A5eihauRQvo

https://www.youtube.com/watch?v=tovrpoUkzYU&t=89s

https://www.youtube.com/watch?v=X9UhB953TDA