**Using RL to Sustain High-density Populations in Conway’s Game of Life**

**I. Prezentare Generala**

In acest articol autorii au introdus un agent de reinforcement learning in Conway’s Game of Life cu obiectivul de a sustine populatii cu viata lunga si densitate mare. Au experimentat cu o varianta de a subdiviza mediul, alaturi de metoda tubulara standard. Au descoperit ca agentul creeaza structuri statice, iar pentru grile de dimensiuni mici rata de invatare este rapida. In final, s-a abordat problema cu deep Q-network cu rezultate similare.

**II. Metode de Realizare**

S-a introdus agent un agent de reinforcement learning care interactioneaza cu mediul, care este o grila n × n cu margini conditionate de regulile jocului. Spatiul de stari este #S = 2n×n. Spatiul de stari este #A = n 2 . Reward-ul este definit ca densitatea de celule vii in fiecare generatie dupa evolutie.

Jocul se poate termina daca nu mai exista celule vii sau numarul de generatii a trecut de limita setata.

**Algoritmi:**

Autorii au folosit Q-Learning, un algoritm de reinforcement learning care poate fii aplicat sarcinilor secventiale. Este un algoritm off-policy deoarece Q-Learning utilizeaza o politica de comportament, care este de natura exploratorie, pentru a invata politica care maximizeaza valoarea asteptata a recompensei totale porning de la starea actuala.

**Metode de aplicare:**

1. **Q-Learning pe grile de dimensiuni mici,** In aceasta metoda, agentulului i se permite sa actioneze asupra oricarei celule a grilei, cu pana la 50 de celule.
2. **Q-Learning pe o grila mare subdivizata.** În aceasta metoda, o grila mai mare se împarte în regiuni de dimensiuni mai mici, compatibile cu Q-Learning tabular. Un dezavantaj ar fi ca pentru agent regiunile nu sunt conectate și evolutia este nedeterminista.
3. **Retele neuronale convolutionale adanci. Q-Learning.** Reteaua consta dintr-un strat convolutionar, urmat de unul sau doua straturi ascunse dense. Tranzitiile sunt colectate într-un buffer. Pentru calcularea pierderii se folosește o rețea profunda ținta.

**Analiza statistica:**

Toate metodele au fost testate cu o stare initiala fixa și una aleatorie. Pentru a evalua performanța metodelor s-au făcut grafuri cu reward-ul ca funcție de evoluție a jocului. Pentru a se înțelege comportamentul agentului s-a vizualizat jocul.

S-au inspirat din lucrările lui Schulman și Seiden pentru implementarea lor propusa de entropie. Un sistem care începe populat aleatoriu, sub regulile de propagare ale jocului. O modalitate de a măsura creșterea sau derularea în sistem este prin utilizarea unei mărimi analoge cu entropia în mecanica statistică. Ideea generală este să subdivizează sistemul și măsoară măsura în care celulele vii se grupează sau nu.

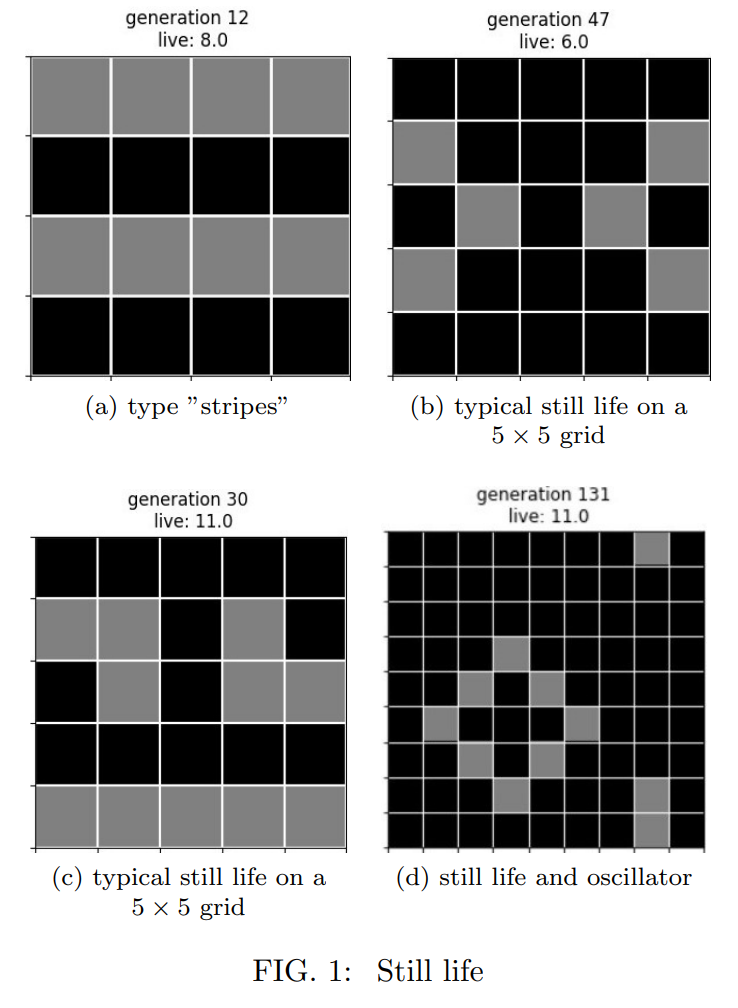
**III. Rezultate**

**Q-Learning pe grile de dimensiuni mici:**

Modelul a fost antrenat pe grile de dimensiune 7 x 7, iar comportamentul a fost identic pentru stari initiale fixe și aleatorii.

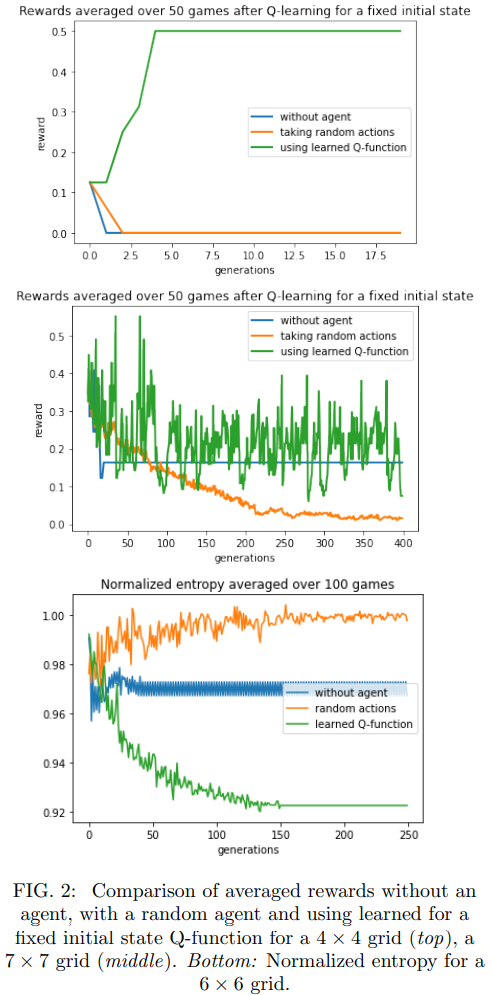
Pentru grile de mărime 4 x 4 rata de invatare este rapidă, iar soluția optima din diverse stari initiale sunt vieți statice de tip dungi.

Pentru grile de mărime 5 x 5 și 6 x 6 continua sa apara vieți statice de diverse forme. Apar și câțiva oscilatori.



Pentru grile de dimensiune 7 x 7 rareori se observa viața statica. Agentul sustine jocul pe perioade lungi de timp. Evolutia e non-periodica, alternand intre stari de densitati mari și mici.

A fost comparata entropia normalizata a starilor create de agentul jocului cu entropia evolutiei naturale și a actiunilor aleatorii. Luând acțiuni complet aleatorii șterge ordinea care apare. De aceea entropia cu acțiuni aleatorii se duce rapid la unul. Contrar la acest comportament, jocurile cu agent tind să aibă un nivel scăzut de entropie. Viața statica este prin natura sa structura ordonata. Cu toate acestea, chiar și atunci când agentul nu poate conduce jocul la o viața statica creează încă structuri de celule care sunt foarte corelate.

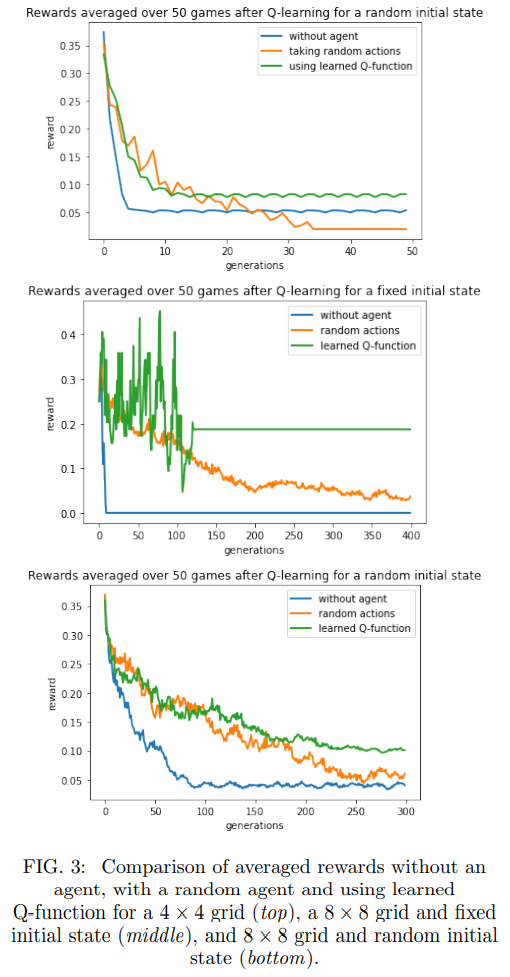


**Q-Learning pe o grila subdivizata:**

Metoda este rulata pentru dimensiuni de subregiune de până la 8 × 8 . Când este antrenat pe întregul spaţiu de stat agentul se comportă mai rău decât evoluția acțiunii aleatoare, asemănătoare cu cea naturală evoluţie. Antrenamentul cu un seed fix duce la un comportament ușor mai bun în primii pași, progresând identic cu evoluția aleatoare.

**Retele neuronale convolutionale adanci. Q-Learning:**

Pentru un 4 × 4 grilă și seed fix rezultatele sunt aceleași cu cele cu metode tabelare. Modelul converge rapid spre viata statica de tip dungi. Când se antrenează cu o stare inițială aleatorie observăm în continuare aceeași soluție, alte forme de viața statica apar ocazional. Cu toate acestea, multe jocuri mor rapid, ceea ce aduce media în jos și performanța generală nu este la fel de bună ca cea corespunzătoare metode tabelare. Pentru grile mai mari cu seed fix, învățarea vieții statica este realizat, dar după mai mulți pași. Învăţarea este semnificativ mai lenta fără seed fix. Dupa antrenament pentru un timp fezabil, apare doar o ușoară îmbunătățire față de acțiuni aleatorii.



**IV. Concluzii**

Autorii au constatat că metoda tabelară standard este robustă pentru această problemă, dar limitat. Metoda tabelara cu subdiviziuni nu a obținut rezultate mai bune decât cea standard, positbil din cauza nedeterministului adaugat. Iar Q-learning a reușit să obținem rezultate bune, dar învăţarea nu a fost foarte stabilă și nu a putut îmbunătăți metodele tabelare.