Atari Breakout - deep reinforcement learning

Aleksandar Nedaković
Fakultet tehničkih nauka - Novi Sad

Uvod

Primarni problem ovog projekta je obučiti agenta da igra Atari Breakout igru i da za cilj ima da maksimizuje skor. Skor se dobija uspešnim razbijanjem bloka, tako što se pogodi lopticom. Lopticom se gađaju blokovi, tako što se ona udari platformom, čije kretanje kontroliše agent. Agent ukupno ima 5 života i gubi jedan u trenutku kada loptica padne na tlo pored platforme.
Za simulaciju igre, koristila se OpenAI Gym biblioteka, a sam projekat je rađen u Python jeziku.

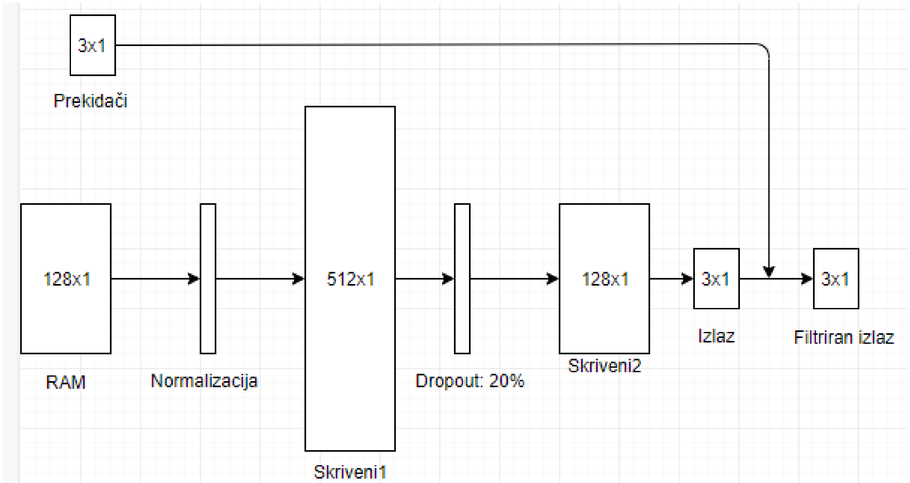
Rešenje problema

Da bismo naučili agenta da postigne visok skor u igri, korišćen je Deep Reinforcement Learning (DRL). Ovo znači da koristimo neuronsku mrežu koja kao ulaz dobija trenutno stanje igre, a kao izlaz daje optimalan potez za odigrati u trenutnom stanju. Kao predstavu trenutnog stanja igre, OpenAI Gym okruženje nam nudi dve opcije:

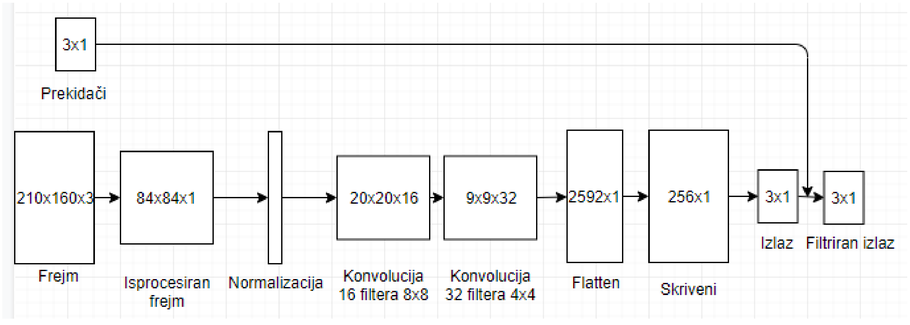
- trenutno stanje 128 bajta koji predstavljaju radnu memoriju Atari konzole
- trenutno prikazani frejm igrice dimenzija 210x160x3

U projektu je vršeno treniranje na obe vrste prikaza, sa tim da se korišćena neuronska mreža znatno razlikuje između prikaza. Kod RAM prikaza, koristi se potpuno povezana neurnoska mreža (*Slika 1.*), dok kod frejm prikaza, frejm se prvo predprocesira na dimenziju 80x80x1 i zatim provlači kroz konovlutivnu mrežu pa na kraju kroz potpuno povezanu (*Slika 2.*).
Da bismo znali koji je optimalan potez za trenutno stanje, izlaz iz mreže čine tri broja koja odgovaraju svakoj od tri moguće akcije (ne radi ništa, pomeri platformu u levo, pomeri platformu u desno) i čija vrednost označava koliko je ta akcija "dobra" za izvršiti. U određenom stanju. činimo akciju čiji izlaz iz mreže je najveći, ali da bismo agenta naterali da istražuje, za određenu verovatnoću (exploration rate) biramo nasumičnu akciju za izvršiti (ta verovatnoća opada vremenom, jer na početku treniranja je potrebno da više istražujemo nego kasnije).
Takođe, na ulazu u mrežu, pored samog stanja imamo i tri "prekidača" za svaku od akcija na izlazu iz mreže. Ako je vrednost na "prekidaču" 0, izlaz za tu akciju je 0, a ako je 1, izlaz odgovara vrednosti dobijenoj u mreži. Ovo se koristi prilikom treniranja, jer naš agent u jednom stanju može da izvrši samo jednu akcjiu i tada palimo prekidač samo na akciji koju je izvršio da backpropagation ne bi uticao na druge dve akcije koje nismo izvršili u tom stanju.

Izgled mreža



Slika 1. Neuronska mreža sa 128 bajta RAM-a na ulazu



Slika 2. Neuronska mreža sa frejmom igre na ulazu

Vreme treniranja

Treniranje mreže je prvenstveno uslovljeno koeficijentom istraživanja (exploration rate). U ovom projektu se vrednost tog koeficijenta linearno smanjivala od 1 do 0.1 tokom prvih 1000000 frejmova igranja, što je vremenski trajalo oko 30 minuta. Tada, agent već ima prosečan skor oko 20 i od tada se skor sporije povećava. Nakon oko 10 sati treniranja, skor polako prestaje da se povećava i dalje treniranje ne dovodi do napretka.

Rezultati

Rezultati se malo razlikuju između korišćenih mreža. Performanse dobijene korišćenjem potpuno povezane mreže i stanja RAM-a kao ulaza su bolje od onih dobijenih konvolutivnom mrežom sa frejmom kao ulazom. Posle par sati treniranja, RAM model je imao prosečan skor oko 30, dok je konvolutivni model imao oko 22. Jednom prilikom je RAM model treniran oko 12 sati bez prestanka i tada su postignuti najbolji rezultati: prosečan skor oko 50, a rekordni skor u nekim partijama je bio čak 350.

```
Game number: 6000    Avg score:33.01    Frame:2655453
Game number: 6100    Avg score:28.33    Frame:2725893
Game number: 6200    Avg score:27.53    Frame:2793148
Game number: 6300    Avg score:30.87    Frame:2866250
Game number: 6400    Avg score:30.25    Frame:2939707
Game number: 6500    Avg score:28.98    Frame:3009715
Game number: 6600    Avg score:32.7     Frame:3081888
Game number: 6700    Avg score:33.75    Frame:3157245
Game number: 6800    Avg score:27.84    Frame:3227109
```

Slika 3. Rezultati sa RAM-om na ulazu

```
Game number: 5900    Avg score:23.12    Frame:2447180
Game number: 6000    Avg score:23.29    Frame:2510306
Game number: 6100    Avg score:22.44    Frame:2568449
Game number: 6200    Avg score:21.54    Frame:2628026
Game number: 6300    Avg score:21.94    Frame:2689627
Game number: 6400    Avg score:20.91    Frame:2747516
Game number: 6500    Avg score:21.59    Frame:2806136
Game number: 6600    Avg score:22.03    Frame:2866151
Game number: 6700    Avg score:23.03    Frame:2929987
```

Slika 4. Rezultati sa frejmom igre na ulazu

Korišćene reference

- Mnih, Volodymyr, et al. "Playing atari with deep reinforcement learning." arXiv preprint arXiv:1312.5602 (2013).
- https://becominghuman.ai/lets-build-an-atari-ai-part-1-dqn-df57e8ff3b26
- https://becominghuman.ai/beat-atari-with-deep-reinforcement-learning-part-2-dqn-improvements-d3563f665a2c