Obučavanje RL agenta za Chrome Dinosaur Game

Q-Learning, Deep Q-Learning, Neuroevolucioni genetski algoritam

Filip Ivković – SW33/2016

Uvod

Chrome Dinosaur Game je popularna igra ugrađena u Google Chrome browser. Pravila igre su vrlo jednostavna – agent (*dinosaurus)* kreće se kroz okruženje nailazeći na prepreke. Dozvoljene akcije su skok i čučanj. Cilj igre je prevazići što veći broj prepreka i samim time postići što veći skor. Igra se prekida u trenutku kolizije agenta i prepreke.

Cilj istraživanja

Cilj istraživanja je obučavanje *reinforcement learning* agenta koji igra Chrome Dinosaur Game implementacijom algoritama koji bi agenta doveli do inteligentnog ponašanja u vidu prevazilaženja prepreka i maksimizovanja ukupnog skora.

Prikaz agenta i okruženja

Nakon implementacije, treniranja agenta i testiranja, moguće je doneti zaključak koji algoritmi donose najbolje rezultate za ovaj problem, i koji su razlozi iza toga.

Algoritmi

Agent vrši interakciju sa okruženjem u realnom vremenu. U svakom koraku agent vrši observaciju okruženja. Na osnovu toga, agent bira optimalnu akciju koju izvršava, prelazeći u novo stanje, i od okruženja dobija nagradu. Cilj agenta je da maksimizuje kumulativnu sumu nagrada kroz vreme. Algoritmi korišćeni prilikom rešavanja problema su:

Q-Learning

- Najosnovniji RL pristup - u toku interakcije sa okruženjem agent unapređuje procenu vrednosti stanja i beleži ih u tabeli.

Deep Q-Learning

- Duboka konvoluciona neuronska mreža
- Kao ulaz prima sliku trenutnog stanja igre. Nad slikom se vrši preprocesiranje kako bi se povećala efikasnost algoritma.

Neuroevolucioni genetski algoritam

- Jednostavna feed-forward neuronska mreža čije su težine inicijalizovane na nasumičan način
- Kroz evolucioni process, genetski algoritam optimizuje težine neuronske mreže kroz process selekcije, ukrštanja, mutacije i reprodukcije najboljih jedinki u svakoj generaciji.

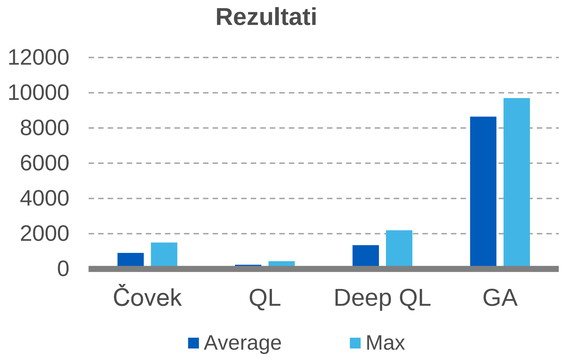
Kao izlaz, svi algoritmi generišu optimalnu akciju za stanje sa ulaza. Moguće akcije su skok, čučanj i trčanje.

Rezultati

U tabeli je dato poređenje prosečnih i maksimalnih skorova prosečnog ljudskog igrača i algoritama korišćenih za obučavanje agenta.

	Average	Max
Čovek	910	1500
Q-Learning	234	440
Deep Q-Learning	1347	2196
GA	8647	9696

**Poređenje je vršeno na nivou 20 igara*



Za Q-Learning, navedeni rezultati dobijeni su nakon 1200 iteracija, za Deep Q-Learning nakon 3000 iteracija, a za genetski algoritam je testiranje izvršeno nakon 40 generacija, gde je veličina populacije 1000.

Zaključak

Upotrebom pomenutih algoritama za obučavanje, agent dolazi do inteligentnog ponašanja koristeći isključivo vizuelnu reprezentaciju igre i nagrada koje mu daje okruženje. Na osnovu rezultata testiranja, može se doneti sledeći zaključak za svaki od algoritama:

- **Q-Learning:** Sporo konvergira iz razloga što sva stanja čuva u tabeli i posmatra ih nezavisno. Neupotrebljiv za promenljivu brzinu kretanja agenta, daje validne rezultate samo za konstantnu brzinu kretanja.
- **Deep Q-Learning:** Treniranje traje relativno dugo ali dovodi agenta do ponašanja koje prevazilazi sposobnosti prosečnog ljudskog igrača. Ograničenje ovog pristupa je što je što agent teže istražuje i uči o stanjima u kojima je brzina kretanja velika, iz razloga što svaka nasumčna akcija koja podstiče eksploraciju dovodi do kraja igre.
- **Neuroevolucioni GA:** Veoma brzo dovodi do ponašanja agenta koje čovek ne može da reprodukuje zbog ograničenja čula vida i sporog vremena percepcije i reakcije. Za konkretni problem se pokazao kao najefikasniji pristup, I u vremena treniranja, i u vidu dobijenih rezultata. Razlog iza toga je što je primarni zadatak genetskog algoritma optimizacija vrednosti, za razliku od dubokih neuronskih mreža koje se fokusiraju na prepoznavanje šablona.

Departman za računarstvo i automatiku
Fakultet tehničkih nauka, Novi Sad
ftn.uns.ac.rs