

TD-Gammon implementacija igre Backgammon

Ana Mitrović, Nikola Vidič

Matematički fakultet

18.09.2017

O igri

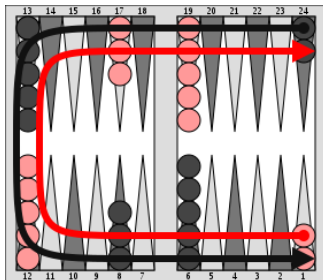
Tavla ili bekgemon je jedna od najstarijih igara na ploči i igrala se još u Persijskom carstvu, drevnom Egiptu i Rimskom carstvu. U ovoj igri na tabli učestvuju **dva igrača**. Svaki od igrača ima po **15 žetona** koji se pomeraju preko **24 polja** (trouglastog oblika) nakon bacanja kockice. Cilj ove igre je biti prvi čiji su žetoni uklonjeni sa table.



Početne pozicije i osnovni pojmovi

Svaka strana ploče ima po 12 šiljastih trouglova koji se numerišu od 1 do 24 u smeru potkovice. Polja od 1 do 6 nazivaju se home board ili inner board, a od 7. do 12. polja outer board.

Kretanje se vrši od većih brojeva prema manjima i cilj je da svih 15 figura dođe u svoj home board.



Kretanje i napadanje protivničkih figura

Na početku igre, svako od igrača baca po jednu kocku i onaj igrač koji je dobio veći broj odigrava dva dobijena broja. Nakon toga, u svakom sledećem bacanju, igrači naizmenično bacaju po dve kocke i igraju dobijene brojeve. U slučaju dobijenih istih brojeva na kockama, na primer 3 i 3, figure se pomeraju za dobijeni broj četiri puta.

Ako se na određenom polju nalaze dve ili više protivničkih figura, to polje je zauzeto (odnosno, protivničke figure na tom polju su nejestive) i igrač ne može na njega da stavi svoju figuru.

Kretanje i napadanje protivničkih figura

Ako se na polju nalazi samo jedna protivnička figura, igrač može pojesti protivničku figuru koja tada privremeno ide izvan igre, u bar (sredina ploče). Time je igrač protivnika usporio i vratio mu jednu figuru skroz na početak, jer pre daljeg pomicanja svojih figura, on mora prvo da izvadi svoju figuru iz bara. Ako je polje zauzeto s dve ili više protivničke figure, na to mesto nije moguće izaći iz bara.

Izbacivanje figura i kraj igre

Pre nego što počne da izbacuje figure s ploče, igrač je dužan svih svojih 15 figura da dovede u home board. Tek tada može početi da izbacuje figure na način da sa svakim brojem kocke izbacuje figure koje su na polju tog broja.

U slučaju da igrač na kockama dobije broj kojim ne može direktno izbaciti nijednu figuru s table, dužan je odigrati figure s pozicije na većem broju. Ako su veće pozicije prazne, onda s ploče vadi figuru s prvog manjeg broja. (Ako je pozicija 4 prazna, a igrač dobije broj 4, mora odigrati figuru koja se nalazi ili na poziciji 6 ili na 5. Ako su i te dve pozicije prazne, tada vadi s ploče figuru koja se nalazi na broju 3)

Pojam

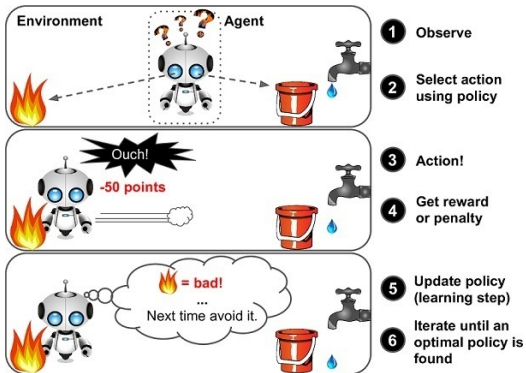
TD-Gammon (Temporal Difference) je neuronska mreža koja samu sebe trenira i kao rezultat dobija ciljnu funkciju igre. Mreža se trenira igrajući protiv sebe i učeći na osnovu rezultata izlaza.

Reinforcement learning

Metoda mašinskog učenja koja se zasniva na učenju na osnovu iskustva. Program koji uči razmatra ulaz, izvršava akciju kojom proizvodi izlaz, a zatim dobija od okruženja nagradu (*reinforcement*) koja mu govori koliko je dobru ili lošu odluku doneo. Cilj je pronaći akciju koja vodi ka najvećoj nagradi tako da ukupna suma nagrada bude maksimalna.

Reinforcement learning

Agent istražuje okruženje u kome se nalazi i cilj je da nakon određenog vremena "nauči" koje radnje su dobre a koje loše.



Temporal credit assignment

Čest je slučaj da je odgovor okruženja (nagrada) odložen, npr. nagrada sledi nakon niza ulaza/izlaza. Tada, program koji uči treba da reši tzv. problem privremene dodele nagrade (*temporal credit assignment*) - program treba da shvati kako da svakom paru ulaz/izlaz dodeli nagradu ili kaznu.

Osnovna ideja TD metoda

Ocena akcije u trenutku t treba da odgovara oceni u trenutku $t+1$. Drugim rečima, ako program kaže da je određeni potez dobar, u narednom koraku treba da se pokaže da je potez zaista bio dobar. Recimo, na primeru sa slike stanja koja su agenta dovela do toga da se opeče treba da dobiju manju ocenu tj. kaznu, a ne samo stanje koje je prethodilo tome da se opeče.

Hvala