Projektni zadatak ❋ Strojno učenje

Prirodoslovno-matematički fakultet

Sveučilište u Zagrebu

Pametne crtice

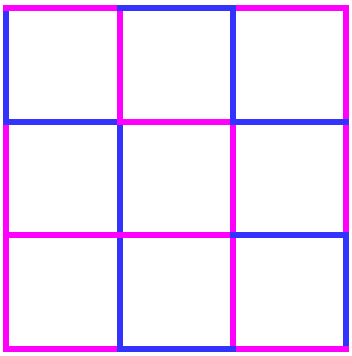
Dario Jurić

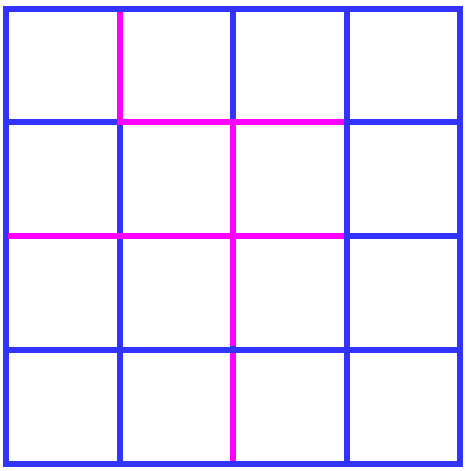
Računarstvo i matematika

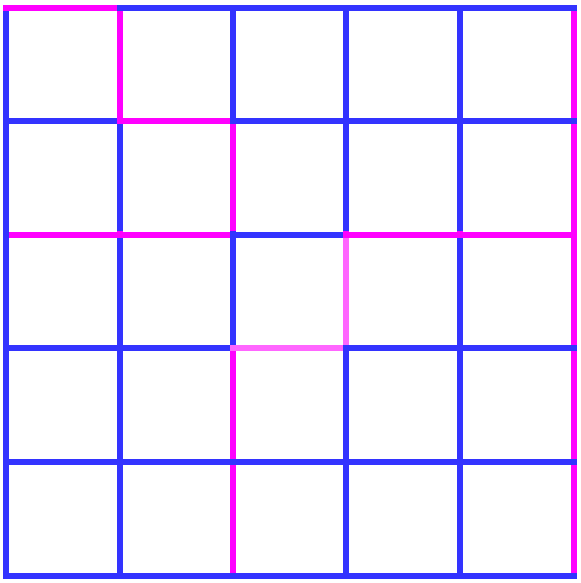
Diplomski studij

Uvodni opis problema

Pametne crtice predstavljaju igru za dva igrača na kvadratoj ploči dimenzija mxn ($m,n \in \mathbb{N}$). Najčešće se igra odvija na ploči dimenzija 3x3, 4x4 ili 5x5:







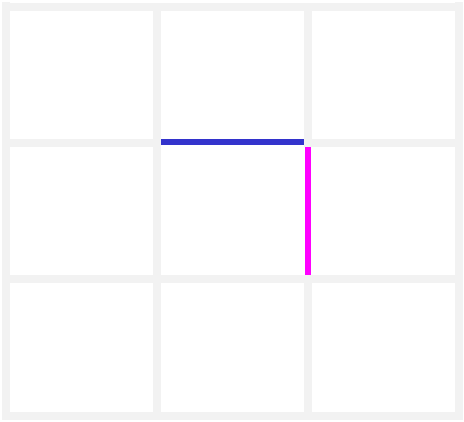
Igra se događa u konačnom vremenskom intervalu $[t_1, t_n]$ $t_i \in \mathbb{Q}$ $1 \leq i \leq n$ $n \in \mathbb{N}$
i vrijedi: $t_1 < t_2 < .. < t_n$. Pojedina crtica na ploči određena je sa parom prirodnih brojeva (m, n) $m,n \in [1,m]_{\mathbb{N}} \text{x} [1,n]_{\mathbb{N}}$ odnosno sa skupom cijelih brojeva $\{a,b\}$.

Princip aktivnosti

Prvi potez

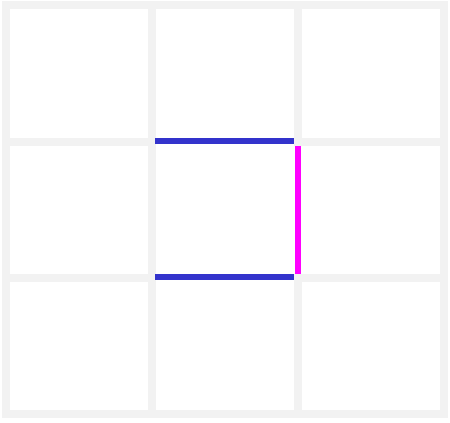
U vremenu t_1 igrač A povlači horizontalnu ili vertikalnu liniju na nekom od slobodnih mjesta na ploči:

Nakon njega, u vremenu $t_2 > t_1$ igrač B po istom principu povlači liniju na nekom od preostalih slobodnih mjesta na ploči:

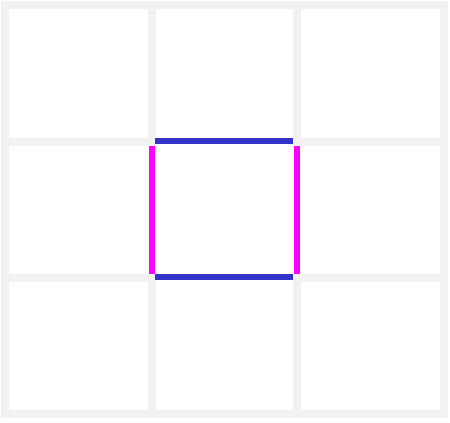


Sljedeći potezi

Zatim, u vremenu $t_3 > t_2$ igrač A kreira liniju na nekom od preostalih slobodnih mjesta na ploči:



Nakon njega, sa ciljem da na ploči napravi kvadrat igrač B u vremenu $t_4 > t_3$ ostvaruje svoj 2. potez:



U ovoj igri, nakon 2. poteza igrač B je ostvario 1 bod.

Igra se odvija na ovaj način dok god ima slobodnih crtica na ploči.

Imamo: svijet $S = [1, m]_{\mathbb{N}} \times [1, n]_{\mathbb{N}}$, dva agenta: Igrač A i Igrač B i skup mogućih akcija $A \subseteq S$.

Projektni zadatak

Pomoću Strojnog učenja uraditi Igrača A koji bi znao igrati Pametne crtice protiv Igrača B koji bi bio neki Čovjek ili računalni program koji slučajno odabire slobodna mjesta na ploči.

Cilj i hipoteze istraživanja problema

Realizirati Igrača A tehnikom Strojnog učenja koji bi se mogao takmičiti protiv Igrača B.
Hipoteza je da će Igrač A biti u stanju pobijeđivati Igrača B.

Materijali, metodologija i plan istraživanja

Projektni zadatak želim riješiti pomoću Strojnog učenja, metoda učenje podrškom (Reinforcement learning), konkretno pomoću algoritma Q učenja (Q-learning)
Agent A := Q-learning agent.
Q learning algoritam i cjelokupni Projekt sam ostvario pomoću JAVA sučelja (u prilogu).
U njemu sam realizirao Svijet gdje se događaju stanja i akcije, Q Agent, Slučajnog agenta, sučelje za komuniciranje sa akcijama Agenta Čovjeka i Grafičko sučelje.
Za realizaciju Q learning agenta potrebno je bilo napraviti što je bolje moguće:

Skup mogućih nagrada $R = \{ r_{11}, r_{12} \dots r_{mn} \}$ $(r_{ij} \in \mathbb{Z})$

funkciju nagrade $r: S \times A \to R$

funkciju prijelaza: $\delta: S \times A \to S$

evaluacijsku Q-matricu $Q: S \times A \to R$ $Q(s,a) = r(s,a) + \gamma V^*(\delta(s,a))$

sa ciljem ostvarenja najbolje optimalne strategije $\pi^*: S \to R$ $\pi^*(s) = \arg\max_a (Q(s,a))$.

Uspješnost realizacije svog projekta mislim vrednovati pomoću rezultata igara Q learning agent protiv Čovjeka, pomoću rezultata igara Q learning agenta protiv Slučajnog agenta i naravno pomoću mišljenja Profesora i Asistenata o kvaliteti Projekta.

Način realizacije projektnog zadatka

Projektni zadatak sam realizirao Java programskim jezikom pomoću Q-learning tehnologije.

Svijet. Stanja. Akcije

Svijet u kome se problem rješava sastoji se od vektora sa m·n stanja ~ cijelih brojeva [1, m·n] i na osnovu njih događaju se moguće akcije koje su opisane skupom cijelih brojeva $\{a, b\} \in [1, m \cdot n] \times [1, m \cdot n]$, tako da se sve opisuje dinamički pomoću ploče koja se sastoji od vektora dimezije maksimalno mn mogućih stanja, te skupa LN koji je skup ostvarenih akcija (ostvarenih linija) kojih može biti $m \cdot (n-1)+(m-1) \cdot n$.
U igri se može izraditi najviše $(m-1) \cdot (n-1)$ kvadrata.
Igra se događa razmjenom poteza (akcijama) dvaju agenata, u multiagent okruženju.

Agenti

U projektnom zadatku se nalaze realizirana 3 IT agenta:

Q-learning agent

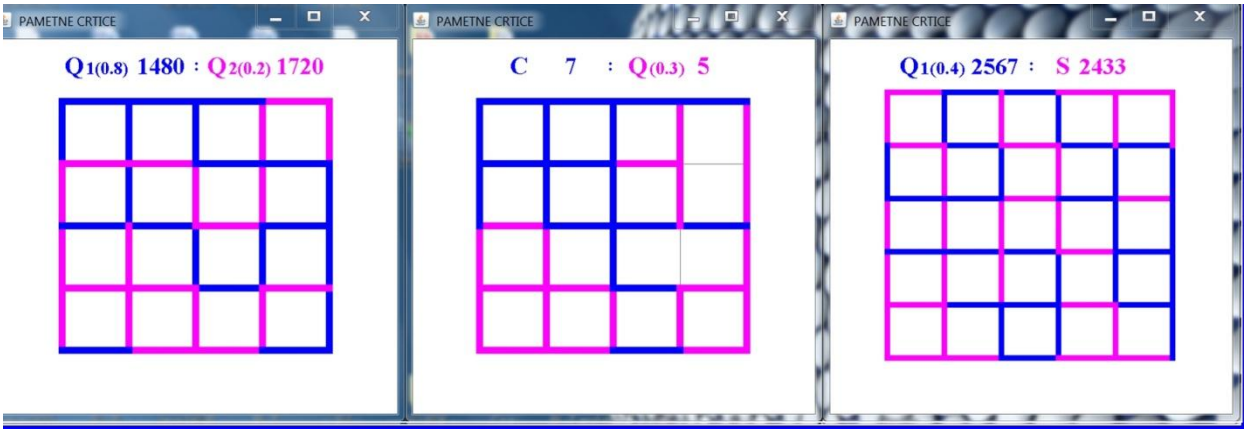
Q-learning agent je realiziran po Q-learning algoritmu:

> Započni sa Q-matricom jednakom nuli.
>
> Izaberi neko moguće stanje $s_t \in [1, m \cdot n]_{\mathbb{Z}}$
>
> Na osnovu njega izaberi sljedeće stanje s_n
>
> Na osnovu funkcije nagrade r(s, a), svih akcija a iz s_n, Q-matrice Q(s,a), $\gamma \in \langle 0, 1 \rangle_{\mathbb{R}}$ izračunaj novu vrijednost Q-matrice Q(s,a) po formuli:
>
> $$Q(s,a) = r(s,a) + \gamma \, Q(s, a \in A)$$
>
> Uzmi za s_t stanje sn trenutno obavljene akcije.
>
> Ponavljaj ovo do pobjede.

Ovime je opisana jedna epizoda Q-learning agenta.
Njih može biti tisuće i Q-learning agent se različito ponaša ovisno o parametru $\gamma \in \langle 0, 1 \rangle_{\mathbb{R}}$
Njime je opisano da li Q-learning agent više cijeni trenutnu nagradu R(s,a) ili prethodna iskustva zapisana u matrici Q(s,a). Obzirom da se do pobjede dolazi kroz trenutne nagrade u analizi tijekom mnogih epizoda pokazalo se da su bolji agenti sa parametrom γ koji je bliži 0, odnosno koji stavljaju naglasak na trenutnu nagradu.

Agent Čovjek

U Projektnom zadatku realizirano je i sučelje za čovjeka kao suigrača Q-learning agenta. Agent čovjek može biti bilo koji čovjek. Pametne crtice realiziraju se po pravilima igre pomoću unosa koordinata cijelih brojeva za jedan potez, npr: 7 8. Nastoji uočiti moguće situacije u kojima može ostvariti bodove. Zatim ponavljaj ovo do pobjede.
Pokazalo se je Q-lerning agent u ovom Projektnom zadatku realiziran kao jedan od ravnopravnih suparnika Čovjeku u većini epizoda.



Slučajni agent

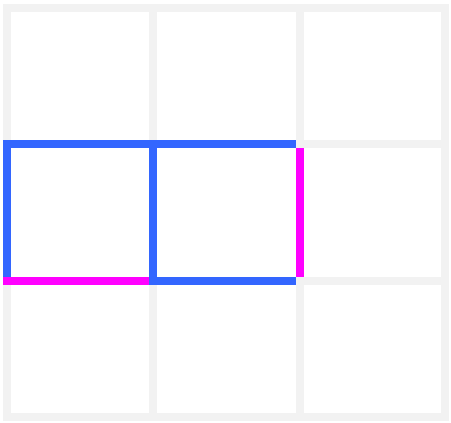
Slučajni agent je realiziran po algoritmu:

> Izaberi neko moguće stanje $s_t \in [1, m \cdot n]_{\mathbb{Z}}$
> Na osnovu stanja st izaberi sljedeće moguće stanje s_n (akciju).
> Na osnovu ostvarene akcije zabilježi ostvarene bodove.
> Ponavljaj ovo do pobjede.

Slučajni agent je realiziran kao jedan od mogućih suparnika Q-learning agentu za analizu istog i pokazao se slabijim u većini epizoda (tisuće njih).
Umjesto njega može se postaviti neki drugi model strojnog učenja radi analize istog.

Moguća poboljšanja

Moguće je poboljšati Q-learning agenta na način da se u dinamičkom okružju svijeta koji se mijenja pretražuje prostor trenutnih nagrada r(s,a) i traži maksimalna moguća.
Naime, moguće su situacije kada se može ostvariti 2 boda tijekom igre:



Prema tome, moguće je realizirati novog agenta, QR Agenta po algoritmu:

Započni sa Q-matricom jednakom nuli.

Izaberi neko moguće stanje $s_t \in [1, m \cdot n]_{\mathbb{Z}}$ na osnovu $\max(R(s \in S, a \in A))$

Na osnovu njega izaberi sljedeće stanje s_n

Na osnovu funkcije nagrade r(s, a), svih akcija a iz s_n,

Q-matrice Q(s,a), $\alpha, \gamma \in \langle 0, 1 \rangle_{\mathbb{R}}$ izračunaj novu vrijednost Q-matrice Q(s,a) po formuli:

$$Q(s,a) = \alpha \max(R(s, a \in A)) + \gamma \max(Q(s, a \in A)) \quad \alpha, \gamma \in \langle 0, 1 \rangle_{\mathbb{R}}$$

Uzmi za s_t stanje s_n trenutno obavljene akcije.

Ponavljaj ovo do pobjede.

Popis literature

Strojno učenje PMF Zagreb

Machine learning Stanford

Machine learning Wikipedia

Machine learning Tom Mitchell

Machine learning AIspace

Dots and boxes MathWorld

Dots and boxes Wikipedia

Java Tutorialspoint

Java Wikipedia