Struktura seminarskog rada

Nacrt naslova rada

Pronalaženje optimalnog poteza u šahu upotrebom dubokog i pojačanog učenja

Autori

Bojan Baškalo

Darko Tica

Publika

- Ko je vaša publika?

  Naša publika su ljudi koje zanima primena mašinskog učenja u igrama, poput šaha, odnosno ljudi koji vole šah i zanimaju ih nova rešenja u polju veštačke inteligencije primenjene na *zero-sum* igre.

- Zašto je publici stalo do teme rada?

  U okviru veštačke inteligencije postoje različiti pristupi rešavanju problema donošenja optimalnog poteza u šahu. Publika želi da sazna koja poboljšanja a koje mane donosi primena dubokog pojačanog učenja (*deep reinforcement learning*) u odnosu na prethodna rešenja.

- Šta publika očekuje da pročita u radu?

  Publika očekuje da pročita sa kojim problemima se može susresti prilikom korišćenja dubokog pojačanog učenja kao i da sazna koje prednosti ovakav metod donosi u šahu.

- Šta publika zna, a šta treba pojasniti u radu?

  Publika je upoznata sa pojmom dubokih neuronskih mreža i pojačanog učenja u aspektima veštačke inteligencije koji se zasnivaju na samostalnom učenju agenata. Publika je takođe upoznata sa *Temporal difference* i *Monte Carlo* metodama za pojačano učenje, koje su važne osnove za pojačano treniranje neuronskih mreža za šah.

  Rad treba da pojasni kako se modeluje stanje okruženja, šta je ulaz a šta izlaz iz neuronske mreže, kako se evaluira trenutno stanje na tabli, kao i kako se bira sledeći optimalni potez iz stabla odlučivanja.

- Kakav stav ima publika prema temi?

  Publika je svesna da su računarski programi koji igraju šah još pre 20 godina uspeli da savladaju čoveka, i da se njihova snaga i veština od tada samo povećavala, te zna da su danas računari nepobedivi u takmičenju sa ljudima. Zbog toga, više ih ne interesuje sposobnost računara da pobedi čoveka, već različiti načini pomoću kojih se može doći do inteligentnih šahovskih programa, ali i beneficije za šah kao igru koje takvi programi mogu doneti.

- Koja pitanja bi čitaoci mogli postaviti?

  Čitaoce bi moglo zanimati koje su sličnosti i razlike između ovog, i rešenja koja već nude savremeni, *state-of-the-art* šahovski programi kao što su *Stockfish*, Komodo i *AlphaZero*. Čitaoci bi takođe mogli imati pitanja koja se tiču izbora načina predstave šahovske table u programu, parametara neuronske mreže, kao i izbora načina obilaženja stabla odluke (prilikom izbora sledećeg poteza).

- Šta želite od publike?

  Želimo da publika shvati prednosti korišćenja dubokog pojačanog učenja u šahu, kao i potencijal za njegovu primenu u ostalim igrama. Takođe želimo da ubedimo publiku da je moguće razviti šahovski program koji igra bolje od ljudi, i to bez pomoći ljudske ekspertize u domenu šaha (velemajstora) u procesu njegovog učenja.

Svrha i motivacija rada

- Šta je naša ključna poruka?

  Agenti trenirani dubokim pojačanim učenjem koji ne zahtevaju domenski tim eksperata iza sebe, donose jednako dobre rezultate kao i računarski programi koji su stvarani na osnovu ekspertskog znanja dugi niz godina prije pojave metoda dubokog učenja u okviru ovog problema.

- Svrha

  Rad obrađuje implementaciju dubokog pojačanog učenja u svrhu kreiranja optimalne evaluacione funkcije koja vrši procenu pozicije igrača u šahu (igrač gubi/vodi) a koja predstavlja ključni deo omogućavanja programa da igra optimalne poteze u šahu (pored stabla odlučivanja).

- Motivacija: zbog čega je ovaj problem bitan za rešavanje?

  Želimo da kreiramo program koji će biti u mogućnosti da pobeđuje ljudske, ali i potencijalno druge računarske programe. Rešavanjem tog problema, moći ćemo da kreiramo program koji će biti sposoban da otkriva nove poteze ali i pozicije u šahu koji ljudima možda ne bi pali na pamet, čime bi se doprinelo znanju ali i razumevanju same igre.

Organizacija rada

1. Apstrakt
   - Jasno ćemo navesti šta je problem koji rešavamo u radu
   - Ukratko ćemo objasniti našu motivaciju za kreiranje ovog rada
   - Prednosti i mane duboko pojačanog učenja u treniranju agenata da igraju šah
   - Trenutna *state-of-the-art* rešenja
   - Ključne reči:
     - Šah
     - *Deep learning, deep neural networks*
     - *Reinforcement learning, temporal difference algorithm*
2. Uvod
   - Jasno navedeni *problem statement* i motivacija
   - Objasnićemo kako funkcionišu inteligentni agenti za igranje šaha, kao i koji su dosadašnji pristupi korišćeni u rešavanju tog problema. Napravićemo osvrt na suštinske razlike između našeg, i trenutno popularnih rešenja.
   - Ukratko ćemo opisati naš pristup - korišćenje dubokih neuronskih mreža i pojačanog učenja u cilju kreiranja inteligentnih agenata za igranje šaha
   - Ukratko o rezultatima samog istraživanja, odnos snage našeg i nekih drugih popularnih agenata (*Stockfish, AlphaZero*)
   - Opis organizacije rada po poglavljima
3. Prednosti dubokog pojačanog učenja nad ranijim rešenjima
   - Opisaćemo neka ranija rešenja, kao i apsekte u kojima naše rešenje predstavlja napredak

        - 3.1. Pristupi bez mašinskog učenja
            - Genetski algoritam
            - Manuelno definisane heurističke funkcije
        - 3.2. Pristupi koji koriste mašinsko učenje
            - Duboke neuronske mreže
            - Duboko pojačano učenje (*temporal differences* i *Monte Carlo* algoritmi)
4. Duboko pojačano učenje
    - 4.1. Interna reprezentacija stanja sistema
        - Opis interne reprezentacije stanja šahovske table u programu
    - 4.2. Evaluacija stanja sistema
        - Opis same evaluacione funkcije u koja opisuje stanje table, odnosno pozicije i broj figura na njoj
    - 4.3. Arhitektura neuronske mreže
        - Opis ulaza i izlaza neuronske mreže
        - Opis same arhitekture neuronske mreže
    - 4.4. Duboko učenje uz pomoć *temporal differences* algoritma
        - Razlike ovog i takođe popularnog *Monte Carlo* pristupa
        - Opis *temporal differences* algoritma i njegova implementacija sa dubokim neuronskim mrežama
    - 4.5. Pretraga stabla odluke
        - Opis mogućih pristupa (*minimax*, *alpha-beta pruning,* neuronske mreže, *probability tree search*)
        - Donošenje odluke o sledećem potezu uz pomoć *alpha-beta pruning*-a
5. Rezultati
    - Opis podataka koji su korišćeni za treniranje mreže (mečevi velemajstora)
    - Rezultati koji su postignuti
    - Poređenje rezultata sa *state-of-the-art* rešenjima
6. Zaključak
    - Prednosti koje je doneo ovaj pristup pri rešavanju problema
    - Ograničenja korišćenja ovog metoda
    - Moguća buduća poboljšanja
7. Literatura
    - Matthew Lai. **Using Deep Reinforcement Learning to Play Chess, 2015.**

      Rad govori o *Giraffe* - šahovskom programu, koji tako što igra protiv prethodne iteracije sebe otkriva znanje specifično za domen, uz minimalno znanje koje mu daje programer

    - Barak Oshri, Nishith Khandwala. **Predicting Moves in Chess using Convolutional Neural Networks, 2015.**

      Rad opisuje rešenje pomoću konvolucionih neuronskih mreža. Koristimo ga zbog procene evaluacione funkcije kao i za tačku poređenja sa našim

pristupom.

- David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, Timothy Lillicrap, Karen Simonyan, Demis Hassabis, **A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play, 2018.**

  Rad opisuje način funkcionisanja *AlphaZero* programa, koji se trenutno smatra najboljim rešenjem kreiranja inteligentnog agenta u igri šaha, i predstavlja okosnicu rešenja u većem broju igara (šah, go, japanski šah).

- JX Wang , Z Kurth-Nelson , D Tirumala, H Soyer, JZ Leibo, R Munos, C Blundell , D Kumaran, M Botvinick, **Learning to reinforcement learn, 2017.**

  Rad istražuje različite pristupe pojačanog učenja, kao i analizu, prednosti i mane pojedinih.