

Primena dubokog Q učenja na automatsko igranje video igara

Student: Nikola Milev Mentor: Mladen Nikolić

Matematički fakultet Katedra za računarstvo i informatiku
Univerzitet u Beogradu
Srbija
08/05/2018

Sadržaj

1	Uvod	2
2	Mašinsko učenje	3
2.1	Vrste mašinskog učenja	4
3	Neuronske mreže	4
4	Konvolutivne neuronske mreže	4
5	Učenje potkrepljivanjem	4
6	DQN	4
7	Detalji implementacije	4

1 Uvod

U maju 1997. godine, Geri Kasparov, tadašnji svetski šampion u šahu, izgubio je meč protiv računarskog sistem pod nazivom "Deep Blue". Skoro dvadeset godina kasnije, program pod nazivom "AlphaGo" pobedio je profesionalnog ljudskog igrača u igri go. Iako su obe igre strateške i igraju se na tabli, između šaha i igre go postoji ogromna razlika. Pravila igre go dosta su jednostavna u odnosu na šah ali je prostor koji opisuje poteze igre go više od 10^{100} puta veći od prostora koji opisuje poteze šaha. Programi koji igraju šah često se zasnivaju na korišćenju stabala pretrage i ovaj pristup jednostavno nije primenljiv na igru go.

Na čemu je onda zasnovan "AlphaGo"? U pitanju je učenje potkrepljivanjem (eng. reinforcement learning). Ovo je vrsta mašinskog učenja koja počiva na sistemu kazne i nagrade. Podrazumeva se da se sistem sastoji od agenta i okruženja u kom agent dela (vrši akcije) i dobija povratnu informaciju o svom učinku. Poput dresiranja psa, nagradama se ohrabruje poželjno ponašanje dok se nepoželjno kažnjava. Cilj jeste ostvariti što veću dugoročnu nagradu. Međutim, agent mora sam kroz istraživanje da shvati kako da dostigne najveću nagradu tako što isprobava različite akcije. Takođe, preduzete akcije mogu da utiču i na nagradu koja se pojavljuje dugo nakon što je sama akcija preduzeta. Ovo zahteva da se uvede pojam buduće nagrade. U učenju potkrepljivanjem, ove dve ideje su ključne.

Pri učenju potkrepljivanjem, najčešće se pretpostavlja da je skup svih mogućih stanja okruženja diskretan. Ovo omogućuje primenu Markovljevih procesa odlučivanja i omogućuje formalan opis problema koji se rešava, kao i pristupa njegovog rešavanja. Formalno predstavljanje problema i rešenja dato je u poglavlju [POGLAVLJE O UCENJU POTKREPLJIVANJEM].

Učenje potkrepljivanjem jedna je od tri vrste mašinskog učenja, pored nadgledanog i nenadgledanog učenja. Pri nadgledanom učenju, sistem dobija skup ulaznih i izlaznih podataka s ciljem da izvrši generalizaciju nad tim podacima i uspešno generiše izlazne podatke od do sada nevidjenih ulaznih podataka. Pri učenju potkrepljivanjem, ne postoje unapred poznate akcije koje treba preduzeti već sistem na osnovu nagrade mora zaključiti koji je optimalni sled akcija. Iako široko korišćeno, nadgledano učenje nije prikladno za učenje iz novih iskustava, kada ciljni rezultati nisu dostupni. Kod nenadgledanog učenja, često je neophodno pronaći neku strukturu u podacima nad kojima se uči bez ikakvog predašnjeg znanja o njima. Iako učenje potkrepljivanjem više liči na nenadgledano učenje, agent ne traži strukturu već teži maksimizaciji nagrade koju dobija od okruženja.

Učenje potkrepljivanjem ima primene u raznim poljima kao što su industrija, istraživanje podataka, mašinsko učenje (kompanije Gugl (eng. Google) koristi učenje potkrepljivanjem radi automatskog dizajna neuralnih mreža), obrazovanje, medicina i finansije. Ovaj vid mašinskog učenja pokazao se kao dobar i za igranje video igara. U radu objavljenom 2015. godine u časopisu "Nature", "DeepMind" predstavlja sistem koji uči da igra video igre sa konzole Atari 2600,

neke čak i daleko bolje od ljudi¹. U avgustu 2017. godine, OpenAI predstavlja agenta koji isključivo kroz igranje igre i bez predašnjeg znanja o igri stiče nivo umeća dovoljan da pobedi i neke od najboljih ljudskih takmičara u video igri "Dota 2"².

U naučnom radu koji je objavila kompanija "DeepMind" u časopisu "Nature" predložen je novi algoritam, DQN (deep Q - network), koji koristi spoj učenja uslovljavanjem i duboke neuronske mreže i uspesno savladava razne igre za Atari 2600 konzolu. Sve informacije dostupne agentu jesu pikseli sa ekrana, kao i trenutni rezultat u igri. Algoritam skladišti prethodna iskustva i umesto učenja neuronske mreže na osnovu samo poslednjih akcija i nagrada, prethodno iskustvo se periodično koristi radi treniranja mreže, smanjujući korelaciju između ulaznih podataka. U sklopu ovog rada, ispitana je struktura algoritma DQN i data je implementacija čije su performanse ispitane na manjoj skali od one date u radu, zbog ograničenih resursa. Takođe je eksperimentisano sa elementima samog algoritma i opisano kako oni utiču na njegovo ponašanje.

[MOZDA NESTO O REZULTATIMA KADA IH BUDE]

U sklopu rada opisani su osnovni pojmovi mašinskog učenja (glava 2), zadržavajući se na neuronskim mrežama uopšte (glava 3) i na konvolutivnim neuronskim mrežama (glava 4). Glava 5 posvećena je učenju potkrepljivanjem dok je algoritam DQN u celosti opisan u glavi 6. U glavi 7 data je implementacija kao i njena evaluacija, dok su eksperimenti i njihovi rezultati opisani u glavi 8.

2 Mašinsko učenje

[izbaci bullshit]

Čovek kroz istoriju teži da posao uradi na najjednostavniji i najlakši način. Ova težnja dovela je do velikih otkrića i preokreta u razvoju civilizacije. Točak je omogućio nastanak kopnenih vozila. Gutenbergova presa ubrzala je umnožavanje knjiga i time ubrzala proširivanje znanja. Otkriće električne struje i njeno eksploatisanje oslobodilo nas je zavisnosti od dnevnog svetla i dovelo do mnogih otkrića bez kojih danas ne možemo da zamislimo život. Jedno od tih otkrića jesu elektronski računari. Idejno, računari su mašine koje mogu da rade prilično jednostavne poslove kao što su osnovne aritmetičke operacije. Daljim razvojem računara, oni su mogli da rešavaju te jednostavne probleme brže. Otkrili smo način da budu manji, jeftiniji i efikasniji i time su postali široko dostupni ali se sama suština računara kao mašine nije promenila. Računari izuzetno brzo rešavaju probleme koje čovek ume formalno da postavi i da algoritam za njegovo rešavanje. Međutim, računari ne umeju da razmišljaju na način na koji čovek razmišlja. Već godinama, naučna fantastika puna je priča o veštačkoj inteligenciji, ljudskoj tvorevini koja parira čoveku po intelektu, mašini koja misli. Ono što danas u računarstvu zovemo veštačkom inteligencijom nije ni blizu tih priča ali se neprestano razvijaju tehnike kojima možemo da implementiramo sistem koji uči da rešava neki relativno uzak opseg problema. Jedna od grana veštačke

¹UBACI NEKU REFERENCU

²<https://blog.openai.com/dota-2/>

inteligencije, mašinsko učenje, bavi se baš korišćenjem statističkih i numeričkih metoda u cilju učenja tj. poboljšavanja uspeha pri obavljanju određenog zadatka.

Probleme koje rešavamo korišćenjem tehnika veštačke inteligencije uglavnom i čovek može da uradi, s tim što bi čoveku to bilo naporno i često bi rešavanje bilo izuzetno sporo. Kada se u problem uvedu apstrakcija i intuicija, nastaje problem. Na primer, prepoznavanje pisanog teksta ili lica na slikama nije nešto što se tek tako može formalizovati. Danas postoje tehnike kojima

2.1 Vrste mašinskog učenja

3 Neuronske mreže

4 Konvolutivne neuronske mreže

5 Učenje potkrepljivanjem

6 DQN

7 Detalji implementacije