
Primena dubokog Q učenja na automatsko igranje video igara

Matematički fakultet
Univerzitet u Beogradu

Student:
Nikola Milev

Mentor:
Mladen Nikolić

Beograd, 2018.

Sadržaj

| | | |
|----------|-------------------------------------|-----------|
| 1 | Uvod | 2 |
| 2 | Mašinsko učenje | 4 |
| 2.1 | Vrste mašinskog učenja | 5 |
| 3 | Neuronske mreže | 6 |
| 4 | Konvolutivne neuronske mreže | 7 |
| 5 | Učenje potkrepljivanjem | 8 |
| 6 | DQN | 9 |
| 7 | Detalji implementacije | 10 |

1

Uvod

U maju 1997. godine, Gari 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.

¹UBACI NEKU REFERENCU

²<https://blog.openai.com/dota-2/>

2

Mašinsko učenje

Mašinsko učenje počelo je da stiče veliku popularnost devedesetih godina prošlog veka zahvaljujući potrebi da se uči iz ogromne količine dostupnih podataka i uspešnosti ovog pristupa u tome. Za popularizaciju mašinskog učenja početkom 21. veka najzaslužnije su neuronske mreže, u toj meri da je mašinsko učenje često poistovećeno s njima. Ovo, naravno, nije tačno; sem neuronskih mreža, postoje razne druge tehnike, kao što su metod potpornih vektora, genetski algoritmi, itd.

Mašinsko učenje nastalo je iz čovekove želje da oponaša prirodne mehanizme učenja kod čoveka i životinja i korišćenja dobijenih rezultata u cilju praktične upotrebe. Termin mašinsko učenje prvi je upotrebio pionir veštačke inteligencije, Artur Semjuel ¹. Doprineo je razvoju veštačke inteligencije istražujući igru dame (eng. checkers) i tražeći način da stvori računarski program koji na osnovu iskustva može da savlada ovu igru. ²

Mašinsko učenje može se definisati kao disciplina koja se bavi izgradnjom prilagodljivih računarskih sistema koji su sposobni da poboljšaju svoje performanse koristeći informacije iz iskustva ³. No, u biti mašinskog učenja leži generalizacija (dedukcija). Dve vrste zaključivanja, indukcija ⁴ i dedukcija ⁵ imaju svoje odgovarajuće discipline u sklopu veštačke inteligencije: mašinsko učenje i automatsko rezonovanje. Kao što se indukcija i dedukcija razlikuju, i mašinsko učenje i automatsko rezonovanje imaju različite oblasti primene. Automatsko rezonovanje zasnovano je na matematičkoj logici i koristi se kada problem čovek relativno lako može formulisati ali ga, često zbog velikog prostora mogućih rešenja, ne može jednostavno rešiti. U ovoj oblasti, neophodno je dobiti apsolutno tačna rešenja, ne dopuštajući nikakav nivo greške. Pri mašinskom učenju, teže je for-

¹https://en.wikipedia.org/wiki/Machine_learning

²<http://infolab.stanford.edu/pub/voy/museum/samuel.html>

³<http://poincare.matf.bg.ac.rs/~janicic/courses/vi.pdf>

⁴Indukcija – zaključivanje od pojedinačnog ka opštem

⁵Dedukcija – zaključivanje od opšteg ka konkretnom

malno definisati problem jer postoji relativno visok nivo apstrakcije. Čovek neke od ovih problema lako rešava a neke ne. Ukoliko je neophodno napraviti sistem koji prepoznaje lica na slikama, kako definisati problem? Od čega se tačno sastoji lice? Metodama automatskog rezonovanja bilo bi nemoguće definisati ovaj problem i rešiti ga. Mašinsko učenje, s druge strane, pokazalo se kao dobar pristup.

Ova oblast je kroz manje od 20 godina od popularizacije postala deo svakodnevnice. U sklopu društvene mreže Fejsbuk (eng. Facebook) implementiran je sistem za prepoznavanje lica koji preporučuje profile osoba koje se nalaze na slikama. Razni veb servisi koriste metode mašinskog učenja radi stvaranja sistema za preporuke (artikala u prodavnicama, video sadržaja na platformama za njihovo gledanje, itd.) i sistema za detekciju prevara. Mnoge firme koje se bave trgovinom na berzi imaju sisteme koji automatski trguju deonicama. U medicini, jedna od primena mašinskog učenja jeste za uspostavljanje dijagnoze. Još neke primene su u marketingu, za procesiranje prirodnih jezika, bezbednost, itd.

2.1 Vrste mašinskog učenja

Mašinsko učenje

3

Neuronske mreže

4

Konvolutivne neuronske mreže

5

Učenje potkrepljivanjem

6

DQN

7

Detalji implementacije