UNIVERZA NA PRIMORSKEM
FAKULTETA ZA MATEMATIKO, NARAVOSLOVJE IN
INFORMACIJSKE TEHNOLOGIJE

Zaključna naloga

Učenje iz interakcije

Learning from interaction

Ime in priimek: Rok Breulj
Študijski program: Računalništvo in informatika
Mentor: doc. dr. Peter Rogelj

Koper, Avgust 2013

Ključna dokumentacijska informacija

Key words documentation

Zahvala

Kazalo

Tabele

Slike

1 Uvod

Ideja učenja iz interakcije z našim okoljem je ena od prvih, ki nam pride na misel, ko razmišljamo o naravi učenja. Ko se dojenček igra, maha z rokami ali gleda naokoli nima izrecnega učitelja, ima pa neposredno senzomotorično povezavo z okoljem. Uporaba te povezave proizvede ogromno informacije o vzrokih in učinkih, o posledicah dejanj in načinih kako doseči cilje. Skozi naše življenje so takšne interakcije nedvoumno velik izvir znanja o našem okolju in samim sebi. Ko se učimo voziti avto ali pogovarjati, se zavedamo kako se okolje odziva na naša dejanja in iščemo način kako vplivati na rezultat z našim vedenjem.

What's in this work.

1.1 Okrepitveno učenje

Okrepitveno učenje (angl. reinforcement learning) je učenje kaj narediti, kako izbirati dejanje, da povečamo številčni nagrajevalni signal. Učencu niso nikoli predstavljena pravilna ali optimalna dejanja kot pri večini oblik strojnega učenja. Katera dejanja prinesejo največjo nagrado mora sam odkriti s poizkušanjem. Skozi interakcijo z okoljem se uči posledic svojih dejanj. V najbolj zanimivih in težavnih primerih imajo dejanja vpliv ne le na takojšnjo nagrado ampak tudi na naslednji položaj in posledične nagrade. Te dve karakteristiki, iskanje s poizkušanjem in zamudne nagrade, so dve najpomembnejši lastnosti okrepitvenega učenja.

Okrepitveno učenje ni definirano s karakterističnimi metodami učenja temveč kot karakterizacija problema učenja. Katerokoli metodo primerno za rešitev problema smatramo kot metodo okrepitvenega učenja. Celoten problem okrepitvenega učenja je predstavljen na strani 9. Osnovna ideja je zajeti najpomembnejše vidike realnega problema s katerim se sooča učenec (angl. learning agent) pri interakciji s svojim okoljem za dosego cilja. Takšen učenec mora imeti nekakšna čutila za pridobivanje informacij o stanju okolja in mora biti sposoben vplivati na to stanje z dejanji. Imeti mora tudi cilj ali pa cilje, ki se nanašajo na stanje okolja. Namen opisa problema je predstaviti te vidike, čutenje, dejanje in cilj, v najenostavnejši obliki brez poenostavljenja.

Okrepitveno učenje se razlikuje od nadzorovanega učenja (angl. supervised learning) v tem, da nima izobraženega zunanjega nadzornika, ki predloži učencu primere in rezultate. Nadzorovano učenje je pomemben tip učenja vendar ni primerno za učenje iz interakcije. V interaktivnih problemih je velikokrat nepraktično pridobiti primere želenega vedenja, ki so pravilni in hkrati predstavljajo vsa stanja v katerih mora učenec delovati. V neznanem okolju, kjer bi si predstavljali, da je učenje najbolj koristno, se mora učenec učiti iz svojih izkušenj [[[[[[[[[[1]]]]]]]]]].

Eden od izzivov okrepitvenega učenja, ki jih ne najdemo v ostalih oblikah strojnega

učenja, je kompromis med raziskovanjem (angl. exploration) in izkoriščanjem (angl. exploitation).

1.2 Primeri

1.3 Elementi okrepitvenega učenja

2 Problem

2.1 Ocenjevanje povratne informacije

2.2 Celoten problem okrepitvenega učenja

3 Rešitve

3.1 Dinamično programiranje

3.2 Predvidevanje - vrednost stanja

3.2.1 Monte Carlo metode

3.2.2 Učenje na podlagi časovne razlike - TD(0)

3.2.3 Združitev metod - TD(λ)

3.3 Krmiljenje - vrednost dejanja

3.3.1 Monte Carlo metode

3.3.2 Učenje na podlagi časovne razlike - TD(0)

3.3.3 Združitev metod - TD(λ)

4 Posploševanje in funkcijska aproksimacija

4.1 Predvidevanje - vrednost stanja

4.2 Krmiljenje - vrednost dejanja

5 Učenje na namizni igri Hex

5.1 Ozadje

5.2 Implementacija

6 Zaključek

7 Literatura

8 Priloge