Control prin învăţare (Learning Control)

Laborator 4: Algoritmi online cu aproximare

Regulament – acelaşi ca şi la laboratorul precedent, cu excepţia termenului limită pentru predarea soluţiei, care este **6 iunie 2019**. Pentru a asigura destul timp pentru procesul de corectare, **termenul limită pentru orice soluţie întârziată este tot 6 iunie 2019**. Studenţii care la această dată nu au toate soluţiile predate nu vor fi eligibili pentru examen.

Introducere

Vom considera din nou problema pendulului inversat, vezi laboratorul 3 pentru o descriere. Codul ce implementează simulatorul pendulului inversat şi formează baza laboratorului poate fi descărcat de pe site-ul cursului, secţiunea "Practical assignments". Dezarhivaţi codul într-un director de pe calculatorul dvs, navigaţi din MATLAB în acest director şi rulaţi scriptul `startup`. Implementarea algoritmilor de planificare online este deja furnizată, iar fişierul `lab4_example` ilustrează cum se pot rula aceşti algoritmi şi cum se pot schimba parametrii lor.

Sistemul este vizualizat în timp ce algoritmii rulează, pentru a avea o impresie în "timp real" a performanţei lor, dar această vizualizare poate fi dezactivată dacă se doreşte ca algoritmii să ruleze mai repede. După ce un algoritm finalizează un experiment, acesta este reprezentat grafic, oferind o imagine de ansamblu a performanţei. Fişierul exemplu arată şi cum se poate afla returnul obţinut de-a lungul traiectoriei şi timpul mediu de execuţie al planificării (media este efectuată de-a lungul paşilor traiectoriei).

Cerinţe

Vom investiga întâi comportamentul algoritmului de *învăţare Q aproximată*. Fiindcă algoritmul depinde de explorare, care este aleatoare, analiza comportamentului său va fi una calitativă.[1] Începeţi prin a rula algoritmul cu setările furnizate.

1. Studiaţi calitativ efectul ratei de învăţare `alpha`, folosind o valoare mică, o valoare aproape sau egală cu 1, şi o valoare intermediară (toate ţinute constante de-a lungul învăţării). **[2p]**
2. Studiaţi calitativ efectul ratei de scădere a eligilităţii `lambda`, folosind valoarea 0, o valoare aproape de 1, şi o valoare intermediară. **[2p]**
3. Studiaţi calitativ efectul ratei de scădere a explorării `explordecay` (probabilitatea de explorare scade la sfârşitul fiecărei traiectorii cu o rată dată de această valoare). **[2p]**

Vom investiga şi algoritmul *LSPI online*.

4. Studiaţi calitativ efectul intervalului K între îmbunătăţirile legii de control, folosind de exemplu 10, 100, 1000 paşi (de notat că în cod se setează parametrul echivalent `dupd`, care este egal cu numărul de paşi K înmulţit cu timpul de eşantionare). Puteţi de exemplu să schimbaţi şi secvenţa de explorare, controlând `explordecay`. **[2p]**

[1] O analiză cantitativă ar necesita rularea fiecărui experiment de un număr mare de ori pentru a obţine rezultate semnificative din punct de vedere statistic; acest lucru nu face parte din cerinţele laboratorului.

Pentru un nivel de încredere mai bun în rezultate, fiecare experiment poate fi repetat de câteva ori (rezultatele depinzând de explorare şi fiind aşadar aleatoare). Pentru aceste experimente multiple, vi se recomandă dezactivarea vizualizării conform instrucţiunilor de mai sus, pentru a obţine rezultatele mai rapid.

Pentru ambii algoritmi, încercaţi să obţineţi o performanţă mai bună decât cea obţinută cu valorile furnizate ale parametrilor. Discutaţi procesul de acordare, rezultatele obţinute, precum şi dificultăţile întâmpinate.