

Control prin învățare (Learning Control)

Laborator 4: Planificarea online

Regulament

Regulament – același ca și la laboratorul precedent, cu excepția termenului limită pentru predarea soluției, care este **joi 1 iunie 2017**. Acesta este și **termenul limită pentru orice soluție întârziată**. Studenții care la această dată nu au toate soluțiile predate nu vor fi eligibili pentru examen.

Introducere

Vom investiga comportamentul a doi algoritmi de planificare: *OPD*, planificarea optimistă pentru sisteme deterministe, și *SOPC*, planificarea optimistă simultană pentru acțiuni continue, a se vedea partea 6 a cursului. Reamintim că planificarea are acces la model și ca atare nu trebuie să învețe.

Vom considera din nou problema pendulului inversat, vezi laboratorul 3 pentru o descriere. În funcția de recompensă:

$$\rho(x, u) = -x^T Q_{\text{rew}} x - R_{\text{rew}} u^2$$

vom alege ponderile $Q_{\text{rew}} = \text{diag}(1, 0.01)$, $R_{\text{rew}} = 0.3$, unde valoarea relativ mare pentru R_{rew} este aleasă în scopul optimizării consumului de energie, pe lângă atingerea stării țintă (originea). Această funcție de recompensă este apoi normalizată în intervalul $[0, 1]$. Factorul de discount este ales $\gamma = 0.85$.

Codul ce implementează simulatorul pendulului inversat și formează baza laboratorului poate fi descărcat de la adresa:

<http://busoniu.net/teaching/ci2017>

secțiunea “Practical assignments”. Dezarhivați codul într-un director de pe calculatorul dvs, navigați din MATLAB în acest director și rulați scriptul `startup`. Implementarea algoritmilor de planificare online este deja furnizată, iar fișierul `lab4_example` ilustrează cum se pot rula acești algoritmi și cum se pot schimba parametrii lor.

Sistemul este vizualizat în timp ce algoritmi rulează, pentru a avea o impresie în “timp real” a performanței lor, dar această vizualizare poate fi dezactivată dacă se dorește ca algoritmi să ruleze mai repede. După ce un algoritm finalizează o traiectorie, aceasta este reprezentată grafic, oferind o imagine de ansamblu a performanței. Fișierul `exemplu` arată și cum se poate afla returnul obținut de-a lungul traiectoriei și timpul mediu de execuție al planificării (media este efectuată de-a lungul pașilor traiectoriei).

1 OPD

Pentru OPD, variați bugetul de expansiuni n alocat algoritmului la fiecare pas, folosind câteva valori între 50 și 300. Înregistrați returnurile și timpii de execuție pentru fiecare valoare a lui n , și reprezentați grafic aceste cantități în funcție de n . [1p]

Discutați influența lui n asupra calității soluției și a timpului de execuție. Analiza algoritmului presupune că timpul de execuție este proporțional cu numărul de expansiuni; este această presupunere confirmată de rezultatele reale? De ce/de ce nu? [2p]

2 SOPC

Pentru SOPC, bugetul n este specificat ca un număr de *tranziții* ale sistemului care pot fi simulate, spre deosebire de numărul de expansiuni în arbore în OPD. Ținând cont că expansiunea unui nod în OPD necesită mai multe tranziții, configurați un set de valori ale lui n pentru SOPC în așa fel încât să folosească același număr de tranziții ca și OPD mai sus. Rulați algoritmul pentru fiecare valoare, și reprezentați grafic returnurile în funcție de n . **[1p]**

Comparați aceste returnuri cu cele obținute de OPD pentru aceleași bugete de tranziții, discutând motivele pentru relația observată. **[1.5p]**

3 Studiul semnalului de comandă

Alegeți o valoare a bugetului pentru care ambii algoritmi obțin soluții bune (un singur balans/swing; de ex. bugetul maxim studiat), și studiați traiectoriile obținute de fiecare dintre algoritmi, concentrându-vă pe forma semnalului de comandă u . Puteți identifica diferențele între semnalele obținute de către algoritmi de planificare, și motivele pentru care acestea apar, ținând cont de natura algoritmilor? **[1p]**

SOPC ar trebui să aibă poată optimiza mai bine energia injectată în sistem pentru a-l controla. Pentru a infirma sau confirma acest lucru, vom studia pentru toate valorile lui n energia de comandă E depusă de-a lungul traiectoriei. Vom presupune că puterea electrică este proporțională cu pătratul voltajului $V^2 = u^2$; reamintiți-vă de asemenea că energia este integrala puterii consumate de-a lungul timpului. Reprezentați grafic E în funcție de n pentru OPD și SOPC. Discutați rezultatul. **[1.5p]**