Domáca úloha č. 8

2-AIN-150, Zima 2018

Termín: 8.1.2019, 23:59, moodle.uniba.sk/fmfi

Skôr ako sa pustíte do riešenia domácej úlohy, oboznámte sa so všeobecnými pokynmi, ktoré sú priložené na konci tohto dokumentu. Riešenia, ktoré odovzdáte, musia byť vaše vlastné. Neopisujte a nesnažte sa nájsť riešenia v literatúre alebo na internete!

Učenie odmenou a trestom

Uvažujte prázdnu cestu (úsečka). Na nej v mieste označenom 0 sa nachádza semafor. Máte auto na pozícií a (môže byť aj záporná) a aktuálne má rýchlosť 0. Chcete sa čo najrýchlejšie dostať do pozície 23, tam zastať a neprejsť na červenú.

Túto úlohu budeme riešiť učením odmenou a trestom nasledovne. Stav sa bude skladať z troch zložiek:

- Pozícia auta – celé číslo z rozsahu $-50, 50$
- Rýchlosť auta – celé číslo z rozsahu $-5, 5$ (záporná rýchlosť znamená, že ideme dozadu).
- Farba svetla na semafóre a jeho trvanie – číslo z množiny $\{-5, -4, -3, -2, -1, 1, 2, 3, 4, 5\}$. Kladné číslo f znamená, že zelená bude ešte $f - 1$ ťahov a záporné číslo $-f$ znamená, že červená bude ešte $f - 1$ ťahov.

Množinu stavov tvoria všetky trojice z vyššie uvedených čísel. Stav $(23, 0, x)$ je cieľový dobrý stav (nedá sa tam robiť akcia) a stavy $(50, x, y)$ a $(-50, x, y)$ sú cieľové zlé stavy (nedá sa tam robiť akcia). Akcie môžete urobiť z množiny $A = \{-1, 0, 1\}$. Akcia vám zmení najprv rýchlosť auta (o to číslo aké má akcia) a potom sa podľa rýchlosti zmení pozícia auta. Prechodová pravdepodobnosť je definovaná nasledovne: S pr. 90% vás auto poslúchne a s pr. 5% vykoná každú z ostatných akcií. Odmena je definovaná nasledovne: V dobrom cieľovom stave dostanete odmenu 100. V zlom cieľovom stave dostanete odmenu -100. V ostatných stavoch dostávate odmenu -0.1.

Pomocou vami zvolenej metódy (value iteration, policy iteration, q-learning) naprogramujte funkciu, ktorá vypočíta najlepšiu policy.

Bonusy:

- 3 body – naprogramujte Q-learning a value alebo policy iteration a porovnajte rýchlosť konvergencie.
- 10 bodov – naprogramujte approximáciu Q-funkcie pomocou vhodnej metódy (lineárna regresia, neurónová sieť) a vyhodnotte jej presnosť.

Platí klasika: Nepoužívajte knižnice, ktoré vašu robotu zjednodušia na jeden riadok.

Pokyny pre Python V balíku je súbor `template.py`, v ktorom doprogramujte funkciu `get_policy()`. Vaša funkcia by mala vrátiť najlepšiu policy. V programe máte veľa rôznych pomocných funkciu na zisťovanie nového stavu, simuláciu vášho riešenia, ... Program sa spušťa príkazom `python template.py`.

Pokyny pre iné jazyky Ak chcete úlohu programovať v inom jazyku ozvite sa mi mailom a dohodneme sa.

Všeobecné pokyny

Úlohy odovzdávajte mailom na mail s predmetom uvedeným v nadpise. Svoj kód vložte do prílohy mailu. **Do kódu vložte stručný komentár o vami naprogramovanej metóde.**

Ideálne odovzdávajte domáce úlohy v Pythone (doprogramujte požadované funkcionality zo zadania). Pokiaľ chcete použiť iný jazyk, môžete, ale musíte zároveň naprogramovať aj réžiu okolo (načítanie, výpis, ...). Bolo by ale vhodné, aby som váš program vedel rozbehať pod linuxom bez nutnosti inštalácie komerčných programov (t.j. overte si, či váš Matlabový kód ide spustiť v Octave, C# či funguje pod Monom, ...).