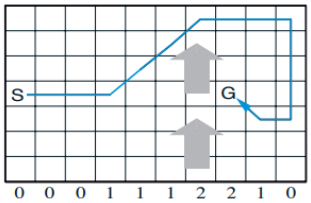
downInteligență Artificială
Laborator 9. Reinforcement Learning

Temă
Considerăm un agent care se poate deplasa într-un mediu (un grid de dimensiuni 7x10). Agentul se poate deplasa în direcția sus, jos, stânga sau dreapta. Punctul de start al agentului, respectiv destinația acestuia sunt prezentate în figura de mai jos prin S (celula (3,0)), respectiv G (celula (3,7)). Există un vânt care poate modifica realizarea acțiunii dorite. În regiunea de mijloc, agentul este deplasat în sus de vânt. Puterea vântului este reprezentată sub fiecare coloană și reprezintă numărul de celule cu care este modificată poziția. Spre exemplu, dacă agentul se află în celula (3,8) și acțiunea este stânga, atunci celula în care ajunge agentul este (4,7). Recompensa este -1 pentru toate tranzițiile. Un episod se termină atunci când agentul atinge obiectivul.



Implementați algoritmul Q-learning pentru a identifica drumul pe care trebuie să-l parcurgă agentul.
a. (0.1p) inițializarea tabelei Q, a parametrilor algoritmului și a stării inițiale
b. (0.1p) pentru o stare s, identifică starea următoare s' prin aplicarea unei acțiuni a
c. (0.7p) algoritmul Q-learning
- selectează acțiunea cu cea mai mare valoare Q din starea s'
- actualizează valorile Q
- actualizează starea curentă
- repetă pașii
d. (0.1p) afișați politica determinată de algoritm

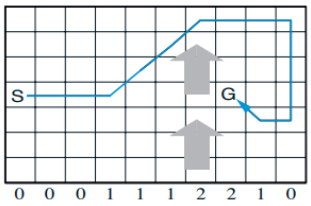
Bonus: (0.1p) verificați convergența algoritmului (spre ex., un grafic ce conține recompensele în raport cu episodul)

Resurse: Secțiunea 6.5 Q-learning: Off-policy TD Control din Reinforcement learning: an introduction http://incompleteideas.net/book/RLbook2020.pdf

Termen limită: laboratorul 10 (4-8 decembrie)
Dacă aveți întrebări legate de temă, puteți trimite un mesaj profesorului cu care faceți laboratorul sau la adresa madalina.raschip@uaic.ro.

Homework
Consider an agent that can move in an environment (a 7x10 grid). The agent can move up, down, left or right. The starting point of the agent, respectively its destination, are shown in the

figure below: S start (cell (3,0)), respectively G goal (cell (3,7)). There is a wind that can modify the resultant next state for a desired action. In the middle region, the agent is moved up by the wind. The wind strength is plotted below each column and represents the number of cells by which the position is changed. For example, if the agent is in cell (3,8) and the action is left, then the next state of the agent will be (4,7). The reward is -1 for all transitions. An episode ends when the agent reaches the goal.



Implement the Q-learning algorithm to identify the path the agent should take.
a. (0.1p) initialization of the Q table, the algorithm parameters and the initial state
b. (0.1p) for a state s, identify the next state s' by applying an action a
c. (0.7p) the Q-learning algorithm
- selects the action with the highest Q value in state s'
- update the Q values
- update the current state
- repeat
d. (0.1p) show the policy determined by the algorithm

Bonus: (0.1p) check the convergence of the algorithm (e.g. a plot containing the rewards over time)

Useful Links: Section 6.5 Q-learning: Off-policy TD Control from Reinforcement learning: an introduction http://incompleteideas.net/book/RLbook2020.pdf

Deadline: lab 10 (December 4-8)
If you have questions regarding the homework, you can send an email to your lab teacher or to madalina.raschip@uaic.ro.