Uczenie ze wzmocnieniem

Marcin Pluciński

mplucinski@wi.zut.edu.pl

Uczenie ze wzmocnieniem – ćwiczenia nr 3

Uczenie ze wzmocnieniem – zadanie 1

- Przedmiotem rozważań będzie gra w zapałki.
- Mamy do dyspozycji pewną ilość N zapałek.
- W grze bierze udział dwóch graczy, którzy na przemian zabierają 1 lub 2 zapałki.
- Przegrywa ten gracz, który zabiera zapałki jako ostatni.

Należy wyznaczyć optymalną strategię gry stosując uczenie ze wzmocnieniem za pomocą algorytmu Q-learning (dla uproszczenia możemy założyć, że nasz przeciwnik będzie grał w sposób losowy).

Uczenie ze wzmocnieniem – zadanie 2

- Przedmiotem rozważań będzie gra w ruletkę.
- Będziemy obstawiać wielokrotności 1 zł na kolor prawdopodobieństwo wygranej wynosi więc $p=\frac{18}{27}$.
- Wyznacz optymalną strategię obstawiania stosując uczenie ze wzmocnieniem, tak by wygrać więcej niż 100 zł.
- Sporządź wykres optymalnej funkcji wartości.

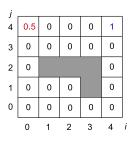
Założenia:

- Stan x określa kapitał gracza w pełnych zł: $x \in \{0, 1, 2, 3, \dots, \}$.
- Możliwe do wykonania akcje (obstawienia) w stanie x: $a \in \{1, ..., x\}$.

Uczenie ze wzmocnieniem – zadanie 3

- Zaprojektuj dwuwymiarowe środowisko komórkowe o wymiarach 10 x 10.
 Podobne środowisko, ale o mniejszym wymiarze analizowane było na pierwszych laboratoriach.
- Rozmieść przeszkody tak, by zajmowały ok. 20% komórek.
- Podobnie jak na prostszym przykładzie pokazanym na kolejnym slajdzie, złóżmy że po osiągnięciu lewego-górnego rogu otrzymamy wzmocnienie równe 0.5, a po osiągnięciu prawego-górnego rogu otrzymamy wzmocnienie równe 1. Stany w tych rogach są absorbujące. Przejścia do pozostałych stanów skutkują wzmocnieniem równym 0.
- Przyjmij epizodyczny tryb uczenia się, przy czym po osiągnięciu stanu absorbującego, nowy stan początkowy powinien być wybierany losowo.
- Zaimplementuj symulator swojego środowiska oraz znajdź optymalną strategię poruszania się w nim. Uczenie zrealizuj z wykorzystaniem algorytmu Q-learning z eksploracją wykorzystującą strategię ε-zachłanną.

Procesy decyzyjne Markowa



- Proces jest deterministyczny.
- Stany: $X = \{(i, j)\}, i, j = 0...4$
- Akcje: $A = \{\uparrow, \downarrow, \leftarrow, \rightarrow\}$
- Stany (0, 4) i (4, 4) sa absorbujace wykonanie dowolnej akcji nie zmienia stanu.

Funkcia wzmocnień:

$$\rho((i,j),a) = \left\{ \begin{array}{l} 1 \; \mathrm{gdy} \; x = (3,4) \; \mathrm{i} \; a = \to \\ 1 \; \mathrm{gdy} \; x = (4,3) \; \mathrm{i} \; a = \uparrow \\ 0.5 \; \mathrm{gdy} \; x = (0,3) \; \mathrm{i} \; a = \uparrow \\ 0.5 \; \mathrm{gdy} \; x = (1,4) \; \mathrm{i} \; a = \leftarrow \\ 0 \; \mathrm{w} \; \mathrm{innych} \; \mathrm{przypadkach}. \end{array} \right.$$

Funkcje przejścia dla pozostałych stanów:

```
\delta((i,j),\uparrow) = (i, \text{ if dozwolone}(j+1): j+1 \text{ else}: j)
\delta((i,j),\downarrow) = (i, \text{ if dozwolone}(j-1): j-1 \text{ else}: j)
\delta((i,j),\leftarrow) = (\text{if dozwolone}(i-1): i-1 \text{ else}: i, j)
\delta((i,j), \rightarrow) = (\text{if dozwolone}(i+1): i+1 \text{ else}: i, j)
```

