Wstęp do sztucznej inteligencji

Uczenie się ze wzmocnieniem

Jakub Robaczewski

Algorytm:

Podstawą algorytmu jest tablica opisująca nagrodę po wykonaniu wybranej akcji zaczynając od wybranego stanu, gdzie kolumny to akcje, a wiersze to stany. Tablica na początku działania jest inicjowana zerami.

Uczenie przebiega przez podaną w zadaniu liczbę ruchów (200), w każdym kroku wybierana jest losowa akcja z podanych. Proces realizowany jest przez metodę learn(), która przyjmuje parametry α i γ oraz liczbę iteracji.

Aktualizacja tablicy przebiega według wzoru:

$$Q_{nowe}(s_t, a_t) = Q(s_t, a_t) + \alpha\left(r_t + \gamma \cdot \max_a Q(s_{t+1}, a) - Q(s_t, a_t)\right)$$

Gdzie:
α – współczynnik uczenia się
γ – współczynnik dyskonta
s – przestrzeń stanów
a – przestrzeń akcji

Wyniki dla trenowania 100 000 iteracji i testowania 10 000 iteracji:

γ \ α	**0.01**	**0.1**	**0.5**	**1**
0.5	0.16	0.08	0.03	0.00
1	0.47	0.84	0.00	0.00
1.5	0.00	0.00	0.00	0.00

Wyniki są najlepsze dla parametrów α = 0.5, γ = 1, jednak najlepsze rezultaty są w granicach 0.70-0.85, co jest spowodowane niedeterminizmem środowiska. Widzimy również, że parametry γ>1 i α>0.5 skrajnie zmniejszają efektywność procesu nauki.