Sieci Samouczące Się

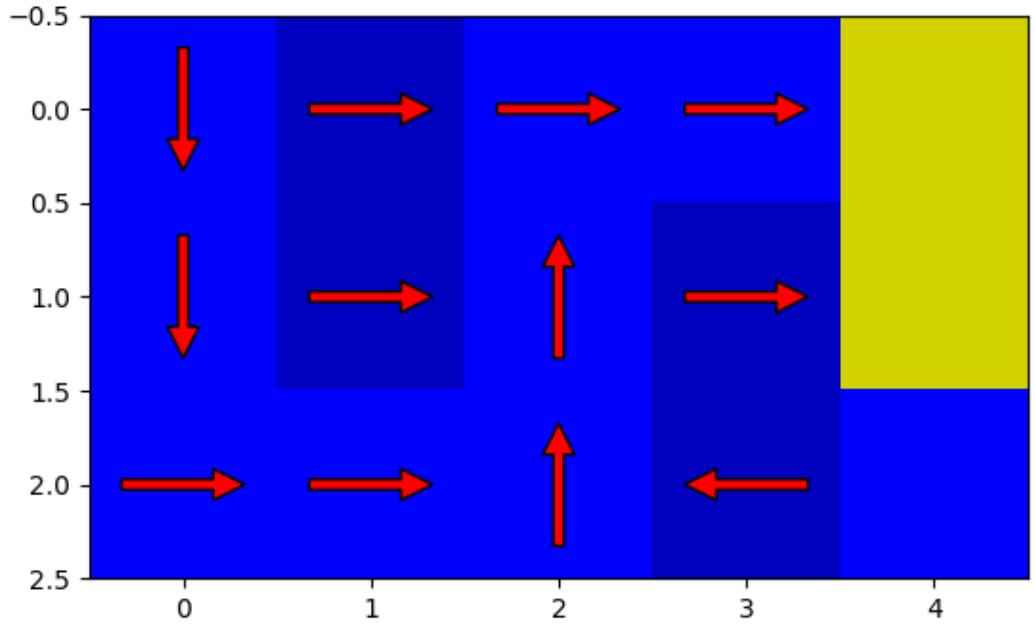
Laboratorium 1 – Sprawozdanie

Metoda Monte Carlo, przyjęty wariant – iteracja wartości, zaimplementowany według slajdów wykładowych.

Działanie – zaczynamy od stworzenia tablicy wartości akcji Q zainicjowanej zerami oraz współczynnika gamma. Następnie dla każdego epizodu ustalamy współczynnik alpha zależny od niego i zaczynamy iterację po możliwych parach (stan, akcja). Dla każdej pary inicjalizujemy stan i akcję początkową oraz tablicę nagród, potem dopóki nie osiągniemy celu lub przekroczymy ustaloną liczbę kroków: wykonujemy akcję w naszym środowisku i zapisujemy jej nagrodę do listy nagród, nadpisujemy stan na nowy stan po wykonaniu akcji i wybieramy metodą zachłanną nową akcję z tablicy Q, która ma aktualnie najwyższą wartość. Po wykonaniu wewnętrznej pętli, aktualizujemy tablicę wartości akcji Q według wzoru, w którym używamy współczynnika dyskontowania gamma.

Wyniki:

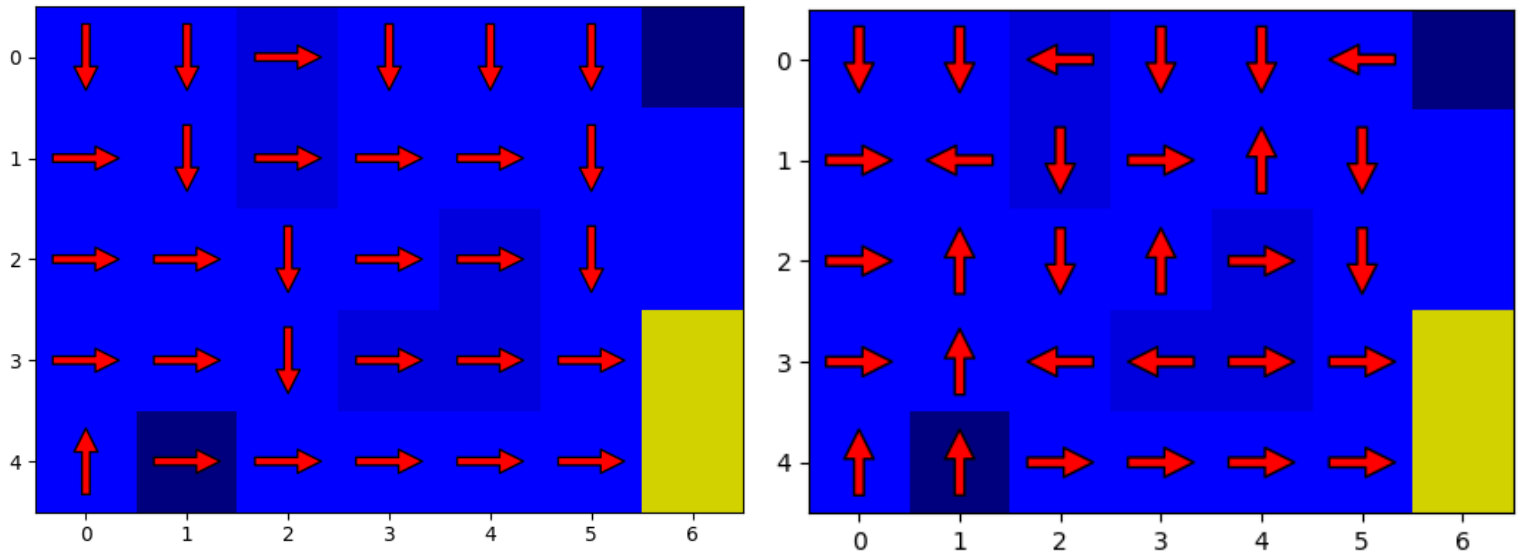
map_small – średnia suma nagród około 4.6 dla współczynnika gamma 0.7 – 1.0, dla niższych wartości wynik się pogarsza



Rys. 1 – wynik dla map_small, 1000 epizodów, gamma równa 0.9

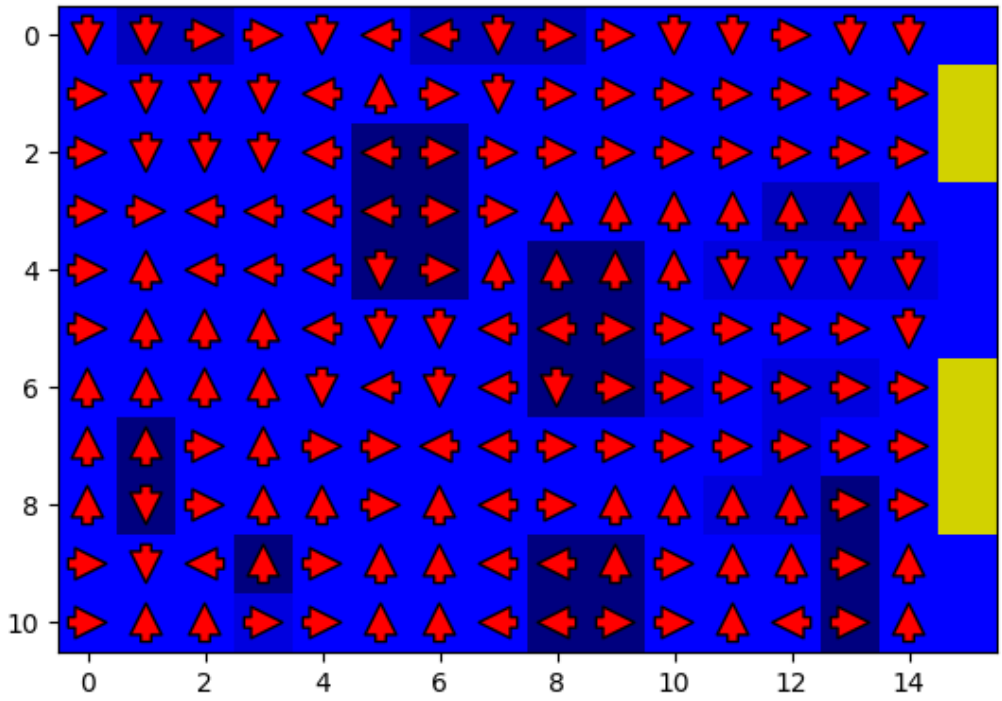
Mateusz Comblik 175827

map_easy – średnia suma nagród w przedziale 6.5 - 7.6 dla gamma 0.6 – 1.0

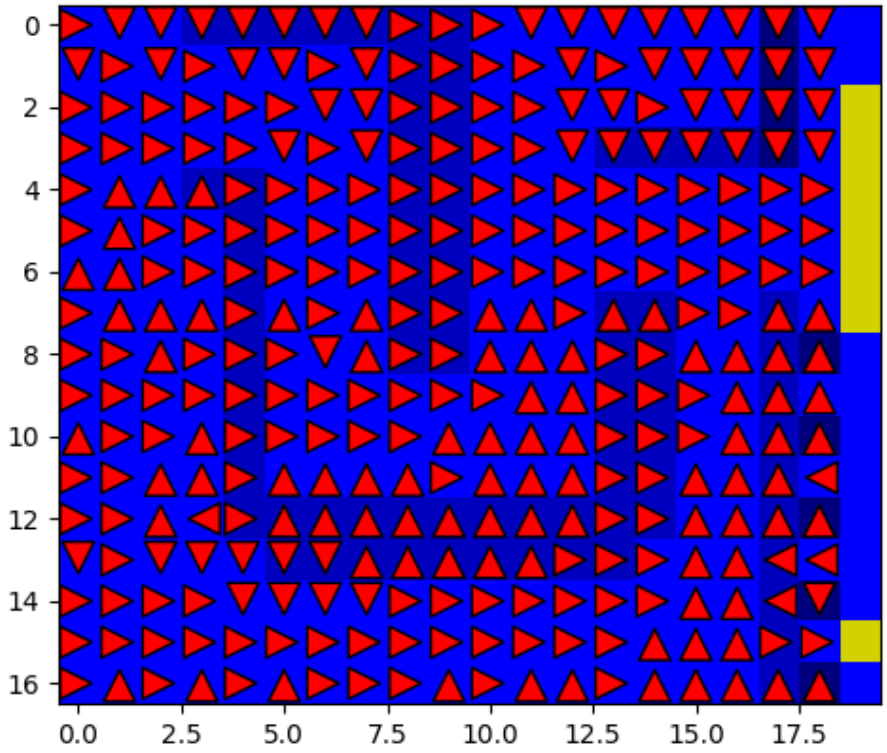


Rys. 2 – wyniki dla map_easy, 1000 epizodów, gamma równa 0.9 po lewej; gamma równa 0.2 po prawej

Widać, że dla mniejszej wartości gamma algorytm staje się krótkowzroczny i może się ‘zapętlać’ zamiast osiągnąć cel.



Rys. 3 – wynik dla map_big, przy gammie 0.9, 1000 epizodów, rewards równe -2.342



Rys. 4 – wynik dla map_spiral, przy gammie 0.9, 300 epizodach, rewards równe 179.64

Złożoność obliczeniowa: z moich testów wynikło, że iteracja strategii jest bardziej złożona obliczeniowo od iteracji wartości.