Sprawozdanie z szóstego laboratorium WSI

Michał Goławski, 325158

1. Opis algorytmu

Algorytm uczy się polityki decyzyjnej dla problemu Frozen Lake 8x8 z włączonym poślizgiem. Algorytm korzysta z Q-learning z epizodami używającego strategii ϵ-zachłannej, która w zależności od parametru epsilon_decay obniża wartość ϵ wykładniczo wraz z kolejnymi epizodami. Algorytm oferuje dwa systemy nagród: domyślny i niestandardowy. Domyślny system nagród zwraca 1 za dojście do celu, 0 za każdą inną akcję. Niestandardowy system nagród zwraca 10 za dojście do celu, -2 za wejście w dziurę, -0.2 za pozostałe akcję, by zachęcić agenta do szybszego zakończenia gry.

2. Opis eksperymentów

Algorytm został przetestowany dla dwóch systemów nagród - domyślnego i niestandardowego. Niezależnie od systemu nagród parametry algorytmu były następujące:

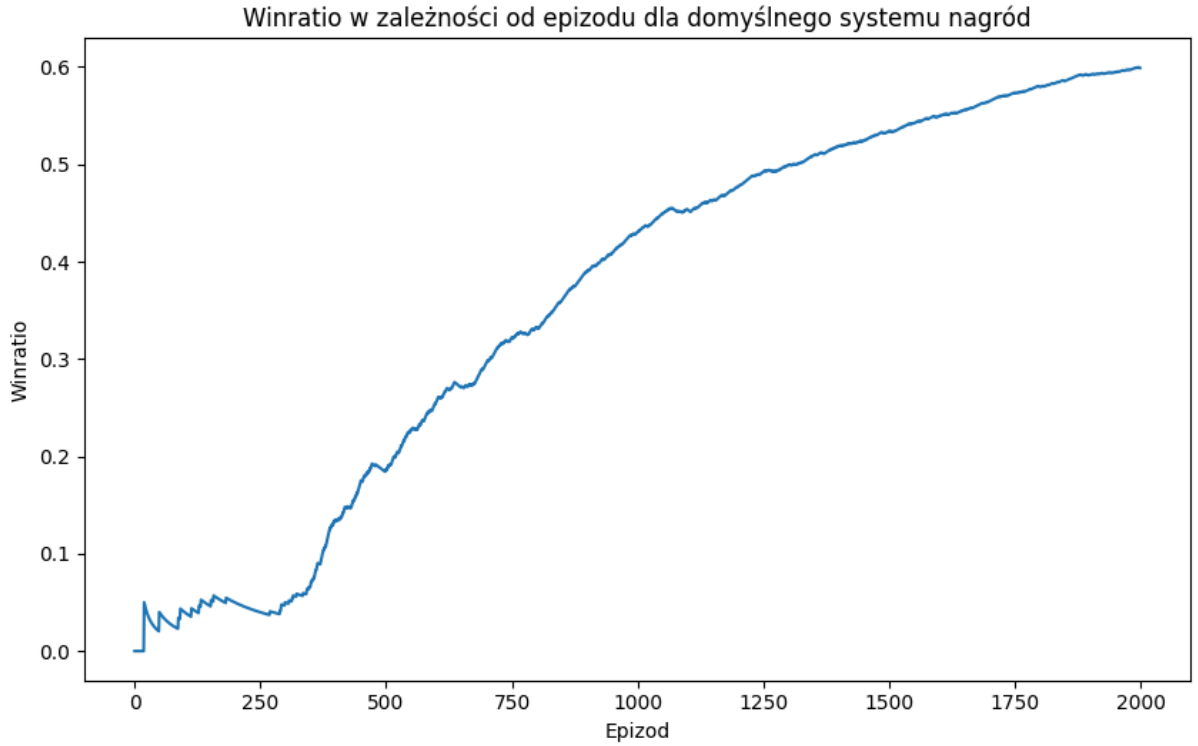
- alfa = 0.9
- gamma = 0.9
- epsilon = 1
- epsilon_decay = 0.01
- episodes = 2000

Eksperymenty polegały na sprawdzeniu w jaki sposób zmieniał się winrate na przestrzeni epizodów podczas treningu w zależności od systemu nagród. Ponadto sprawdzony został winrate algorytmu dla obu systemów nagród na podstawie 200 gier.

3. Wyniki eksperymentów

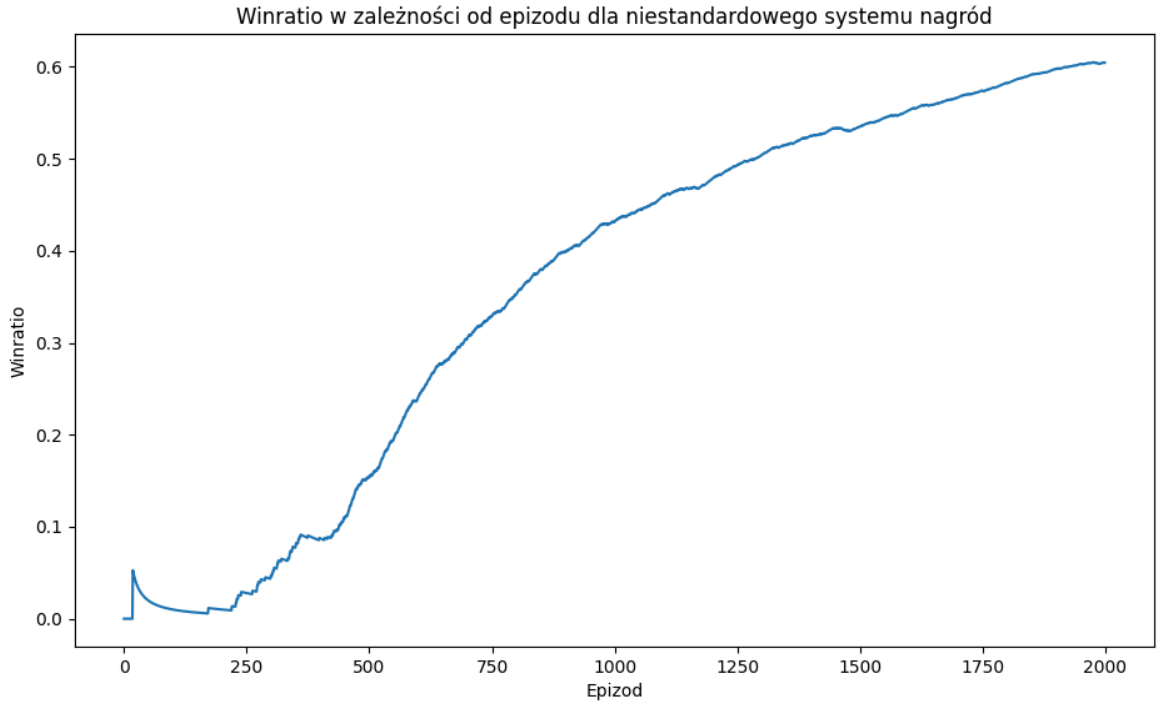
3.1. Domyślny system nagród

Po wytrenowaniu algorytmu z użyciem domyślnego systemu nagród jego winrate na podstawie 200 gier wynosił 77,5%. Tak prezentuje się wykres winrate podczas treningu na przestrzeni epizodów.



3.2. Niestandardowy system nagród

Po wytrenowaniu algorytmu z użyciem domyślnego systemu nagród jego winrate na podstawie 200 gier wynosił 72,5%. Tak prezentuje się wykres winrate podczas treningu na przestrzeni epizodów.



3.3 Wnioski

Jak można zauważyć trening z użyciem domyślnego systemu nagród uzyskał lepsze efekty, jednakże różnica jest stosunkowo mała. Wykresy winratio od epizodów również wyglądają bardzo podobnie do siebie. Warto zauważyć, że jeśli wykona się analogiczne eksperymenty ustawiając parametr epsilon na 0 to niestandardowy system nagród radzi sobie znacznie lepiej. W takim przypadku algorytm wytrenowany na domyślnym systemie nagród osiąga winrate na poziomie ~7% zaś algorytm wytrenowany na niestandardowym systemie nagród osiąga winrate na poziomie ~75%.