

## **Laboratorium 4**

### **Aproksymacja liniowa z kodowaniem metodą pokryć**

#### **Sieci Samouczące Się**

#### **Sprawozdanie**

W celu wykonania laboratorium zmodyfikowałem środowisko uczenia zamieniając wszystko na klasę i funkcje. Klasa Agent() odpowiada za uczenie wahać. Przy tworzeniu klasy są ustawiane zmienne środowiska takie jak maksymalna siła wahać. Do nauki zaimplementowałem algorytm Q-learning z kodowaniem metodą pokryć. Kodowane jest pięć zmiennych [kąt, prędkość kątowa, położenie, prędkość, siła (akcje)], z tego powodu tablica wag również posiada pięć wymiarów. Zastosowałem prosty aproksymator liniowy oraz metodę eksploracji epsilon greedy (1% epsilon to minimalna wartość po spadku). Zmodyfikowałem również funkcję nagrody – dodałem takie zmienne jak kara za zbyt gwałtowną zmianę siły, kara za odległość od celu. Ostateczny algorytm uczy się, jednak odbywa się to bardzo wolno – dopiero po około 3h średnia kroków na 10 epizodów wynosi wartość 200, jednak przy testach często zdarzają się sytuacje gdzie algorytm osiąga kryterium końca 1000 kroków, ale nie zbliża się do środka mapy. Do wytrenowania algorytmu, żeby osiągał oba kryteria – 1000 kroków oraz środek mapy, potrzeba by dłuższego czasu trenowania. Wahać uczy się tak długo z powodu pewnych nieoptymalnych decyzji w kodzie, przede wszystkim przez używanie kodowania pięciu wymiarów, do optymalizacji warto by zastosować mniej wymiarów lub użyć większej ilości aproksymatorów, każdy dla innego wymiaru.