Report

STVIAN004

Ian Stevens

30/10/2014

Description

Q-learning uses reinforcement learning based algorithm. In this assignment we use the minesweeper artificial intelligence modifier to adjust the learning patens of the minesweeper. A learning table is made with the state plotted agents the action and populated with a reward based on the outcome of that action. From this table is can decide on the best course to take in the future.

Method

We make a 2D vector for every minesweeper. The value of this 2D vector is set to 0. For every minesweeper we check to see what action gives us the best outcome base on previous iterations. If they are all the same outcomes for different actions it will pick East – North-West-South.

After every move that outcome of that move is evaluated using the function R(s,a)+ 0.5 \*max(Q(nest state, all actions)). The rewards are -100 for going into a super mine else -1. The -1 is to encourage the mine sweeper to explore new paths. The value is mapped to the chosen action at the position of the mine sweeper and not the cell with a mine. It would be more efficient to just label that cell as bad and not the actions that lead to it ie(don’t have to say at (x,y ) north is bad just say that cell is bad) but I thought this was the proper way

After many iterations the sweepers will learn the best paths to avoid the super mines and hopefully just pick up mines

Pseudocode

Initialise Q table

For each iteration do

For each sweeper do

observe current state (x,y)

select action (direction) with highest historic return

update()

for each sweeper do

observe new state

Update the Q(s,a) using function R(s,a)+ 0.5 \*max(Q(nest state, all actions))