CISC 856 Reinforcement Learning S24

Final Project Proposal

Reinforcement Learning For Connect-4 Game

By Group No. 4

Assem Salama	20482018
Omar Waheed	20481032
Hossam Aboouf	20481027

Queen's School of Computing
2024

Project Description:

Connect-4 is a two players game which takes place on a 7x6 rectangular board placed vertically between them. One player has 21 blue coins and the other 21 red coins. Each player can drop a coin at the top of the board in one of the seven columns; the coin falls down and fills the lower unoccupied square. Of course, a player cannot drop a coin in a certain column if it's already full (i.e. it already contains six coins). The goal of this project is to develop an intelligent agent capable of playing the classic board game Connect-4 using reinforcement learning.

Action Space:

The action space for Connect-4 is relatively small. Each player has at most seven possible actions that they can take at any given state. Therefore, an action is defined as dropping a piece into one of the seven columns on the board. Because the number of actions a player may take at a given state depends on how many columns on the board are full, we calculate the number of actions possible at a given state as:

$$Actions\ Possible = 7 - C \longrightarrow (1)$$

$$C = Number\ of\ full\ columns\ at\ current\ state \longrightarrow (2)$$

State Space:

The state space for Connect-4 is considerably larger than the action space. A state in Connect-4 is defined as the board with played pieces that a player sees. If the player starts the game, then the board will always have an even number of pieces. Alternatively, if they are a second player, the board will always have an odd number of pieces.

A very rough upper bound for the number of possible states of a Connect-4 game can be calculated by considering that each position on the board can either be free, a player one piece or a player two piece. Then because the board size is 6 x 7, we get an upper bound of 3^{42}. However, this calculation does not take into consideration that possible boards are illegal due to gravity and the rules of the game. The best lower bound on the number of possible positions has been calculated by a computer program to be around 1.6×10^{13}.

Proposed Solution:

The targets were calculated according to the Q-learning algorithm (Temporal Difference):

$$Q(s,a) = Q(s,a) + \alpha(max(Q(s',a')) + \gamma R - Q(s,a)) \longrightarrow (3)$$

where Q is the Q function, s is the state, a is the action, α is the learning rate, R is the reward and γ is the discount factor. The reward will be 1 for winning, -1 for losing, 0.5 for a tie and 0 otherwise.