Northeastern

Reinforcement Learning : Tic Tac Toe problem

Kishore Kumar Mohan
Harikrishna Ramakrishna

Tic Tac Toe using Reinforcement Learning

Objective: To play tic tac toe with computer.

To understand how machine understands a tic tac toe problem and learns from it.

Problem Type: Reinforcement Learning Classification

Other problems:

https://www.youtube.com/watch?v=chrzpnL1OEM

Reinforcement Learning:

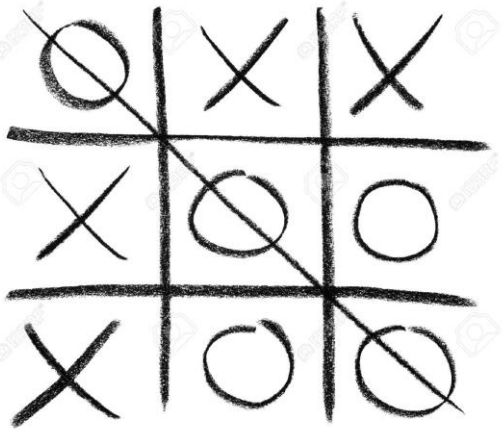
Ideal, human learning occurs either due to one of these,

- Rewards achieved when a task is completed.
- Fear of repercussions when task is incomplete.

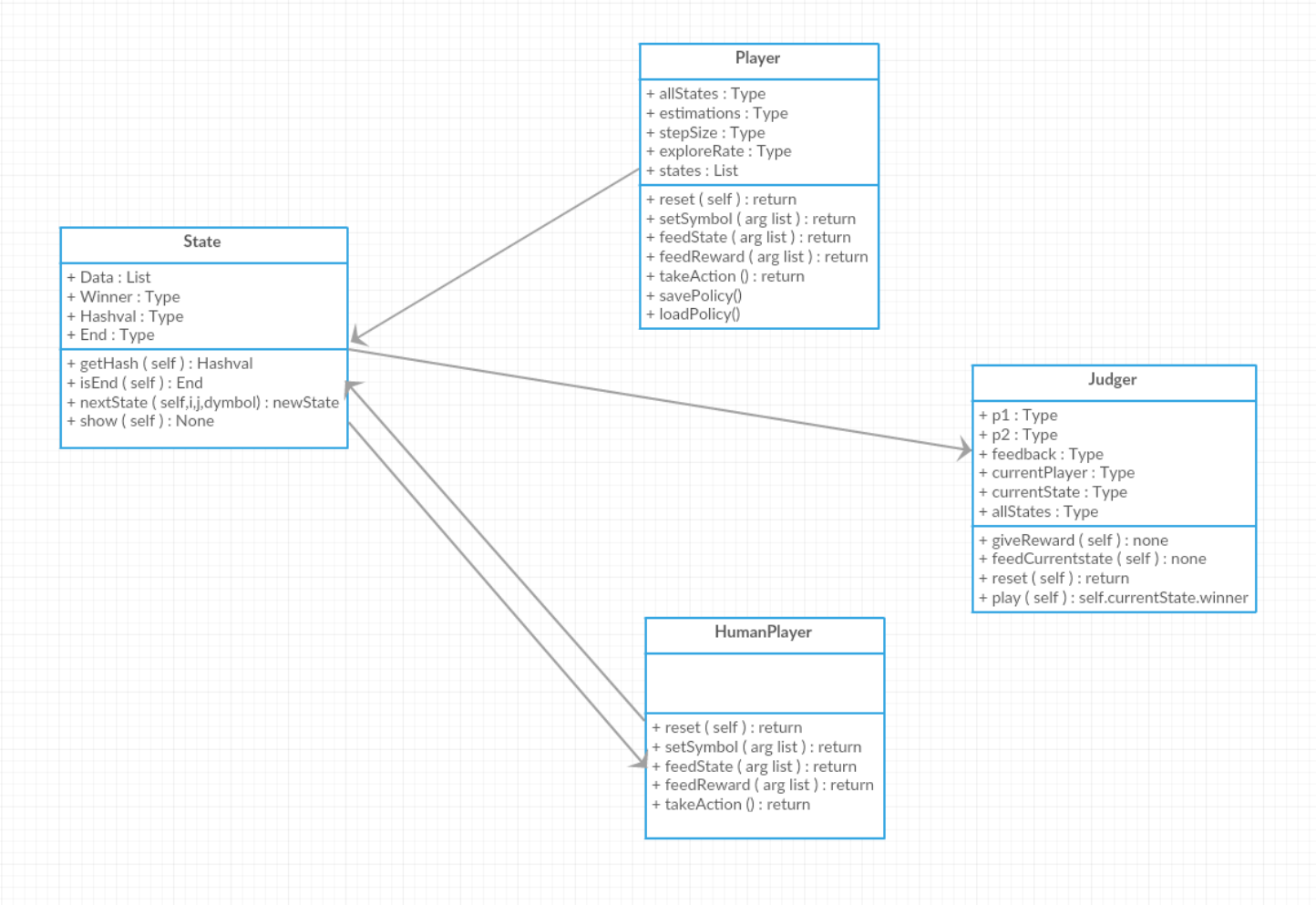
Reinforcement learning occurs when a machine is trained to complete a task with no pre-training steps and yes, assured of a reward for task completion.

T (S,a,S`) - Training  and R(S,a,S`) - Rewards

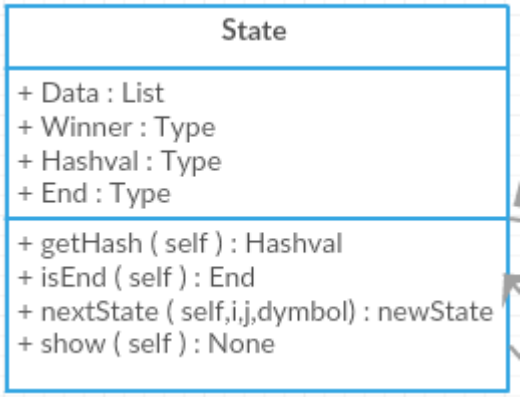
Tic Tac Toe problem:



Tic Tac Toe problem – Class diagram



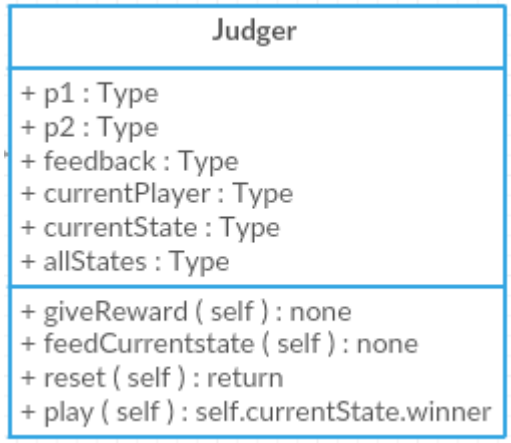
Class - State



What's the purpose of class State?

- State creates data, winner, hashval, end
- data is used to create 3x3 matrix array with all 0s
- Matrix takes values -1,0,1
- If 1, then * and if -1, then x else 0
- getHash() - gives unique value for all 9 boxes
- isEnd() - tie, win, lose condition check
- nextState(i,j,symbol) - puts symbol in specified position
- show() - Displays board

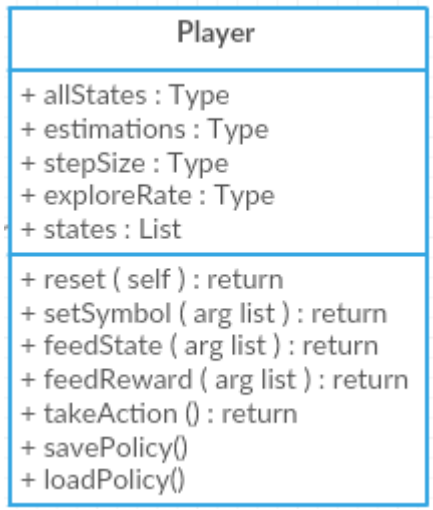
Class - Judger



What's the purpose of class Judger?

- Judger- Initialize players, symbols, rewards and states
- Two players intialization
- Assign and set symbols for each
- Initializing states
- giveReward() - Feedreward for the winning player
- feedCurrentState() - feed moves of each players
- play() - get moves from either players and save it to current state

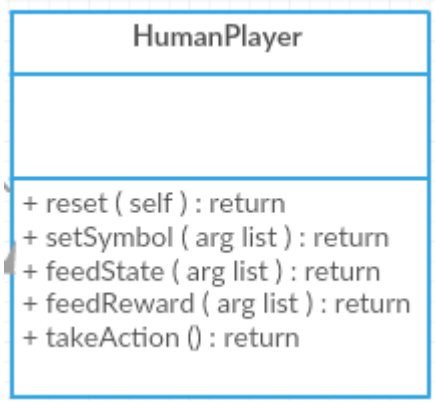
Class – AI Player



What's the purpose of class Player?

- Player - Initialize estimation, stepsize and explorerate
- setSymbol() - save estimations using hashval for future runs. Estimation set to 1 and saved if winner. Else set to 0 and saved if loses.
- takeAction - Using estimations and exploreRate values, take appropriate position to move to next state.
- savePolicy() - saves estimation
- loadPolicy() - loads all saved estimation

Class – HumanPlayer



What's the purpose of HumanPlayer?

- Initializes symbol and currentState
- takeAction() - moves to appropriate position

Main Methods

Method 01: Train

- Initializes two AI players internally and trains them.
- Saves Estimation using savePolicy method.

Method 02: Compete

- Initializes two AI players internally and trains them with explorerate set to 0.0
- Trains player-1 till it learns to win with minimal steps

Method 03: play()

- Challenges human to play.

Summary

- Reading images for a specific architecture
- Why opencv?
- Regular Net – Math and code
- Feeding images after extracting features has zero impact
- Convolutional neural network and activation functions
- Accuracy and Analysis