QLearningAgent Report

Changed numEpisodes to 40,000 (more episodes, for better policy)

Functions Modified:

public void train()

Implemented code to train QLearningAgent for certain number of episodes with the TTTenvironment. For each iteration, the agent played tictactoe against the random agent. Used the epsilon greedy policy to select a move for each state; with a probability of epsilon (0.1), exploring by choosing a random move, otherwise (probability 1-epsilon) exploiting by calculating and choosing the move with highest q-value. After selecting move, executing it in the environment to get sample outcome (s,a,s',r). Then updating the q-value for the state-action pair by also calculating the max q-value of next state (s') and using equation $Q(s, a) \leftarrow (1 - \alpha) * Q(s, a) + \alpha * (r + \gamma * maxQ(s', a'))$. Then updating the q-table for (s,a) and after episode is concluded, environment is reset. In this function at the end the extractPolicy() is called to get optimal policy based on the updated Q-table.

public Policy extractPolicy()

Created a new map to store updated policy values (s,a). Looped through all states in the updated q-table and for each state calculated the q-values for possible moves and selected move with the highest q-value as optimal move. Updated policy with the optimal move for each state.

Testing QLearningAgent

Performance Report

Opponent	Wins	Losses	Draws
Defensive Agent	38	0	12
Aggressive Agent	50	0	0
Random Agent	50	0	0

Results after testing agent in the class file 'TestQLearning.java'

Against Defensive Agent

```
| |O|X|
|X|O|O|
|O|X|X|

Playing move: X(0,0)

|X|O|X|
|X|O|O|
|O|X|X|

It's a draw.
Wins: 38 Losses: 0 Draws: 12
```

Against Aggressive Agent

```
| |O|X|
Playing move: O(0,0)

|O| |X|
| | | |
| |O|X|
Playing move: X(1,2)

|O| |X|
| | |X|
| |O|X|
X won!
Wins: 50 Losses: 0 Draws: 0
```

Against Random Agent

```
|X| |X|
Playing move: O(1,0)

| | |O|
|O| | |
|X| |X|
Playing move: X(2,1)

| | |O|
|O| | |
|X|X|X|
X won!
Wins: 50 Losses: 0 Draws: 0
```

```
Runs: 3/3     Errors: 0     Failures: 0

TestQLearning [Runner: JUnit 4] (5.645 s)
    testAggressive (2.068 s)
    testRandom (1.802 s)
    testDefensive (1.775 s)
```