Blake Johnson

9663980

CS165A Homework 4

1. A. ­ Initial state distribution

{

P({Rock, Rock} = 1/9

P({Rock, Paper} = 1/9

P({Rock, Scissors} = 1/9

P({Paper, Rock} = 1/9

P({Paper, Paper} = 1/9

P({Paper, Scissors} = 1/9

P({Scissors, Rock} = 1/9

P({Scissors, Paper} = 1/9

P({Scissors, Scissors} = 1/9

}

State Transition Matrix

{

P(s’| a, s) = P(s’, a, s)/P(a, s)

P(a, s) = µ(a | s)\*P(s)

P(s’ , a, s) = P(s’ | a, s)\*µ(a | s)\*P(s)

Transition(s’, a, s) = P(s’ | a, s)

Transition(s’, a, s) = P(s’, a, s)/µ(a|s)\*P(s)

}

Reward Distribution

{

State format = {agent, human}

R(s) = 1 for s = {Rock, Scissors},{Scissors, Paper},{Paper, Rock}

R(s) = 0 for s = {Rock, Rock},{Paper, Paper}{Scissors, Scissors}

R(s) = 1 for s = {Scissors, Rock},{Paper, Scissors},{Rock,Paper}

P(R(s’) = 1 | a , s) = P(R(s’) = 1, a, s) / µ(a|s)\*P(s)

P(R(s’) = 0 | a , s) = P(R(s’) = 1, a, s) / µ(a|s)\*P(s)

P(R(s’) = -1 | a , s) = P(R(s’) = 1, a, s) / µ(a|s)\*P(s)

}

B. Q(s) = argmax(a) Σs’ P(s'|s,a)\* U(s')

U(s) = R(s) + γmax P(s'|s,a)U(s')

Q(s) = argmax(a) Σs’ P(s’ | a, s)\* π (a | s)\*P(s)\*U(s')

U(s) = P(R(s,a), a, s) / π (a|s)+ γmax(a) P(s’ | a, s)\* π (a | s)\*P(s)\*U(s')

C. F(π) =

Q(s) = argmax(a) Σs’ P(s’ | a, s)\* π (a | s)\*P(s)\*U(s')

U(s) = R(s) γ max(a) P(s’ | a, s)\* π (a | s)\*P(s)\*U(s')

F(µ) =

Q(s) = argmax(a) Σs’ P(s’ | a, s)\* π (a | s)\*P(s)\*U(s')

U(s) = P(R(s,a) = 1, a, s) / π (a|s)+ γmax(a) P(s’ | a, s)\* π (a | s)\*P(s)\*U(s')

P(R(s,a) = 1, a, s) / π (a|s)+ γ max(a) P(s’ | a, s)\* π (a | s)\*P(s)\*U(s') = P(R(s,a) = 1, a, s) / µ (a|s)+ γ max(a) P(s’ | a, s)\* µ (a | s)\*P(s)\*U(s')

argmax(a) Σs’ P(s’ | a, s)\* π (a | s)\*P(s)\*U(s') = argmax(a) Σs’ P(s’ | a, s)\* µ (a | s)\*P(s)\*U(s')

argmax(a) Σs’ P(s’ | a, s)\* π (a | s)\*P(s) = argmax(a) Σs’ P(s’ | a, s)\* µ (a | s)\*P(s)

argmax(a) Σs’ π (a | s)= argmax(a) Σs’ µ (a | s)

π = µ

When the policies are the same, every next move will be the same until the policy becomes random, which will be the best case scenario and worst case scenario at the

1. A.

B. V\*(s) = maxaΣ Pass’[rass’ + γV\*(s’)]

C. V0π = maxaΣ Pass’[rass’ + γ( maxaΣ Pass’[rass’ + γ( maxaΣ Pass’[rass’ + γ maxaΣ Pass’[rass’ + γVk-2π (s’)Vk-1π (s’)) Vkπ (s’))

V0π(s) = Σk maxaΣs’ Pass’\* rass’ \*(γVkπ (s’))k

D. The value converges to V0π(s) = maxaΣ Pass’(rass’) as k -> inf

1. My translation of the statement is: R -> (C->E) ۸ (⌐C -> ⌐E)

R -> (C -> E) ۸ (C + -E)

R -> (C -> E) ۸ (-E + C)

R -> (C -> E) ۸ (E -> C)

**R -> (C <-> E)**

My reasoning behind this is because the statement is mainly about what being radical implies. This leads to the R -> part of the statement. The rest of the statement comes from the part saying, “if the candidate is conservative then they are electable”, leading to the part (C->E) to the right of the radical implication. The statement then says “but otherwise not electable.” The “but” means ۸, so we and the “otherwise not electable” to the statement. That last part is represented by (⌐C -> ⌐E). When the whole statement, R -> (C->E) ۸ (⌐C -> ⌐E), is reduced, we see that it corresponds to the statement **R -> (C <-> E)**

A. (⌐Food ۷ Party) ۷ (⌐Drinks ۷ Party) -> ((Food ۸ Drinks) -> Party)

(⌐Food ۷ Party) ۷ (⌐Drinks ۷ Party) -> (⌐ (Food ۸ Drinks) ۷ Party)

⌐Food ۷ Party ۷ ⌐Drinks -> ((⌐Food ۷ ⌐Drinks) ۷ Party)

⌐Food ۷ Party ۷ ⌐Drinks -> ⌐Food ۷ ⌐Drinks ۷ Party

B.

|  |  |  |  |
| --- | --- | --- | --- |
| R | C | E | Statement |
| False | False | False | True |
| False | False | True | True |
| False | True | False | True |
| False | True | True | True |
| True | False | False | True |
| True | False | True | True |
| True | True | False | True |
| True | True | True | True |

Valid? Yes, because all values of its variables lead to at least one truthful statement

Satisfiable but not Valid? No, because although its truth table is not a contradiction it is not Valid.

Unsatisfiable? No, because its truth table is not a contradiction