Carissa Ying Geok Teng (A0205190R/E0425113)

1a)

<u>State</u>

On(Start), Path(s_0, s_1) …

<u>Action</u>

Travel(from, to)

Precond: On(from) ^ ~Visited(to) ^ Path(from, to)

Effect: ~On(from) ^ On(to) ^ Visited(to)

<u>Transition</u>

P(s_1|s_0, Travel(s_0, s_1)) …

<u>Reward</u>

R(s) = -0.4

R(Visited(s_0) ^ Visited(s_1) ^ …) = 1

1b)

<u>State</u>

ItemsInStorage(n), NumOfOrders(o), ~Ordered, Backordered(u)

<u>Action</u>

Order(n, m): Order 1 time such that inventory = M

Precond: ItemsInStorage(n) ^ (n + m = M) ^ ~Ordered

Effect: ~ItemsInStorage(n) ^ ItemsInStorage(n + m) ^ Ordered

Backorder(n, u, o): Backorder up to B units

Precond: ItemsInStorage(n) ^ Backordered(u) ^ (n + o < N) ^ (u + o < B)

Effect: ~ItemsInStorage(n) ^ ~Backordered(u) ^ ItemsInStorage(n + o) ^ Backordered(u + o)

FulfillOrder(n, o, f)

Precond: ItemsInStorage(n) ^ NumOfOrders(o) ^ (n – f ≥ 0) ^ (o – f ≥ 0)

Effect: ~ItemsInStorage(n) ^ ~NumOfOrders(o) ^ ItemsInStorage(n – f) ^ NumOfOrders(o – f)

NextDay(n, o, u):

Precond: ItemsInStorage(n) ^ NumOfOrders(o) ^ Backordered(u) ^ (o == 0)

Effect: ~Ordered ^ ~Backordered(u)

<u>Transition</u>

P(New inventory and orders | Old inventory and orders, Action to order more or fulfil orders)

Reward

R(s, Order(n, m)) = - c

R(s, Backorder(n, u, o)) = -b * o

R(s, FulfillOrder(n, o, f)) = f

R(s, NextDay(n, o, u)) = -n

1c)

State

Screen display where each screen pixel with one of its values from 0-127

Action

One of the 18 actions

Transition

P(display after an action is taken| current screen display, one of the 18 actions)

Reward

ΔScore - ΔTime

2a)

Policy

$\pi^*(s_1) = a_2$

$\pi^*(s_2) = a_1$

Value Function

$U^*(s_1) = P(s_1|s_1, a_2)R(s_1|a_2) + P(s_2|s_1, a_2)R(s_2|a_2) = 0.1(0) + 0.9(3) = 2.7$

$U^*(s_2) = P(s_1|s_2, a_1)R(s_1|a_1) + P(s_2|s_2, a_1)R(s_2|a_1) = 0(1) + 1(3) = 3$

2b)

Policy

$\pi^*(s_1, t_1) = a_2$

$\pi^*(s_2, t_1) = a_1$

$\pi^*(s_1, t_2) = a_2$

$\pi^*(s_2, t_2) = a_1$

Value Function

$U^*(s_1) = P(s_1|s_1, a_2)R(s_1|a_2) + P(s_2|s_1, a_2)R(s_2|a_2)$

$+ P(s_1|s_1, a_2)\ P(s_1|s_1, a_2)R(s_1|a_2) + P(s_2|s_1, a_2)\ P(s_1|s_1, a_2)R(s_2|a_2)$ (s_1 on first move)

$+ P(s_1|s_2, a_1)P(s_2|s_1, a_2)R(s_1|a_2) + P(s_2|s_2, a_1)P(s_2|s_1, a_2)R(s_2|a_2)$ (s_2 on first move)

$= 0.1(0) + 0.9(3) + 0.1(0.1)(0) + 0.9(0.1)(3) + 0(0.9)(0) + 1(0.9)(3) = 5.67$

$U^*(s_2) = 3 + 3 = 6$

2c)

Policy

$\pi^*(s_1) = a_2$

$\pi^*(s_2) = a_1$

Value Function

$U^*(s_2) = \frac{3}{1-\gamma} = 3 / 0.1 = 30$

$U^*(s_1) = \gamma P(s_1|s_1, a_2)U^*(s_1) + \gamma P(s_2|s_2, a_2)U^*(s_2)$

$= 0.9(0.1)U^*(s_1) + 0.9(0.9)(30)$

$0.91U^*(s_1) = 24.3$

$U^*(s_1) = 26.703$