(a) Initial settings:
MDP:
S = 1:(1, 1), 2:(1, 2), 3:(1, 3), 4:(2, 1), 5:(2, 2), 6:(2, 3)
A = "N":(-1, 0), "S":(1, 0), "W":(0, -1), "E":(0, 1)
P =
0.8 in desired direction, 0.1 for 2 orthogonal directions to the desired one, stay put if run into the wall

Policy = None (No initial policy)

(b) DFA: (LTL specification: Finally (g3 and g4 and not r5))
$S = [0, 1, 2, 3, 4]$
$Final_{success} = 3$
$Final_{fail} = 4$
$Sink_{states} = [3, 4]$
$A = [g3, g4, r5, phi]$
transitions =
digraph G rankdir = LR
0[label="0",shape=circle]
1[label="1",shape=circle]
2[label="2",shape=circle]
3[label="3",shape=doublecircle]
4[label="4",shape=doublecircle]
2→4[label = "r5"], 2→2[label = "g4"], 4→4[label = "g3"],
3→3[label = "g4"], 0→2[label = "g4"], 4→4[label = "g4"],
1→3[label = "g4"], 1→1[label = "phi"], 1→4[label = "r5"],
0→0[label = "phi"], 3→4[label = "r5"], 3→3[label = "phi"],
4→4[label = "r5"], 3→3[label = "g3"], 2→2[label = "phi"],
2→3[label = "g3"], 0→4[label = "r5"], 1→1[label = "g3"],
4→4[label = "phi"], 0→1[label = "g3"],

(c) product:
$S = cross_{product}(mdp.S, DFA.states)$
$A = mdp.A$
$Final_{cosafe} = cross_{product}(mdp.S, DFA.Final_{success})$
$Final_{fail} = cross_{product}(mdp.S, DFA.Final_{fail})$
$Sink = cross_{product}(mdp.S, DFA.Sink_{states})$
$Label = (2, 1) : g3, (1, 3) : g4, (2, 2) : r5, (1, 1) : phi, (1, 2) : phi, (2, 3) : phi$
P =
As specified in the slides, $P'((s', q')|(s, q), a) = P(s'|s, a)$
only when q' = DFA.transition(Label[s'], q)
Also, set $P(Sink|(s, q), a) = 1$ for all 'action's and all (s,q) in Sink

!Some modifications of initial settings that different from the slides:
(1) Set all rewards to 0, that is, $r((s, q), a) = 0$ for all s, q, and a
(2) Set $V_0[(s, q)] = 1$ for all (s, q) in $Final_{cosafe}$
(3) Set $V_0[(s, q)] = -1$ for all (s, q) in $Final_{fail}$

(4) Set $V_0[(s,q)] = 0$ for all (s,q) not in Sink
(5) Avoid using $P(Sink|(s,q),a)$ by keeping $V_i[(s,q)] = V_0[(s,q)]$ for (s,q) in Sink during the value iteration running process

(d) verification of policy:
(1) gamma = 1.0, threshold = 0.000001, SVI convergence iterations = 15

---

start at: ((1, 1), 0) 0.556177600738 (objective value)
through policy action: S → ((2, 1), 1) 0.558593220739 (objective value) 0.8 chain probability
through policy action: N → ((1, 1), 1) 0.753422368407 (objective value) 0.64 chain probability
through policy action: E → ((1, 2), 1) 0.777777777778 (objective value) 0.512 chain probability
through policy action: E → ((1, 3), 3) 1 (objective value) 0.4096 chain probability
task succeed

---

start at: ((1, 2), 0) 0.53714978801 (objective value)
through policy action: E → ((1, 3), 2) 0.7293664563 (objective value) 0.8 chain probability
through policy action: W → ((1, 2), 2) 0.753422111547 (objective value) 0.64 chain probability
through policy action: W → ((1, 1), 2) 0.972601839551 (objective value) 0.512 chain probability
through policy action: S → ((2, 1), 3) 1 (objective value) 0.4096 chain probability
task succeed

---

start at: ((1, 3), 0) 0.556239614147 (objective value)
through policy action: S → ((2, 3), 0) 0.53714978801 (objective value) 0.8 chain probability
through policy action: N → ((1, 3), 2) 0.7293664563 (objective value) 0.64 chain probability
through policy action: W → ((1, 2), 2) 0.753422111547 (objective value) 0.512 chain probability
through policy action: W → ((1, 1), 2) 0.972601839551 (objective value) 0.4096 chain probability
through policy action: S → ((2, 1), 3) 1 (objective value) 0.32768 chain probability
task succeed

---

start at: ((2, 1), 0) 0.400741703852 (objective value)
through policy action: N → ((1, 1), 0) 0.556177600738 (objective value) 0.8 chain probability
through policy action: S → ((2, 1), 1) 0.558593220739 (objective value) 0.64 chain probability
through policy action: N → ((1, 1), 1) 0.753422368407 (objective value) 0.512 chain probability
through policy action: E → ((1, 2), 1) 0.777777777778 (objective value) 0.4096 chain probability
through policy action: E → ((1, 3), 3) 1 (objective value) 0.32768 chain probability
task succeed

---

start at: ((2, 2), 0) 0.539166787215 (objective value)
through policy action: N → ((1, 2), 0) 0.53714978801 (objective value) 0.8 chain probability
through policy action: E → ((1, 3), 2) 0.7293664563 (objective value) 0.64 chain probability
through policy action: W → ((1, 2), 2) 0.753422111547 (objective value) 0.512 chain probability
through policy action: W → ((1, 1), 2) 0.972601839551 (objective value) 0.4096 chain probability
through policy action: S → ((2, 1), 3) 1 (objective value) 0.32768 chain probability
task succeed

---

start at: ((2, 3), 0) 0.53714978801 (objective value)
through policy action: N → ((1, 3), 2) 0.7293664563 (objective value) 0.8 chain probability
through policy action: W → ((1, 2), 2) 0.753422111547 (objective value) 0.64 chain probability
through policy action: W → ((1, 1), 2) 0.972601839551 (objective value) 0.512 chain probability

through policy action: S → ((2, 1), 3) 1 (objective value) 0.4096 chain probability
task succeed

(e) gamma = 0.9, threshold = 0.000001, SVI convergence iterations = 13

---

start at: ((1, 1), 0) 0.31484740392 (objective value)
through policy action: S → ((2, 1), 1) 0.362877291661 (objective value) 0.8 chain probability
through policy action: N → ((1, 1), 1) 0.583646203241 (objective value) 0.64 chain probability
through policy action: E → ((1, 2), 1) 0.692307692307 (objective value) 0.512 chain probability
through policy action: E → ((1, 3), 3) 1 (objective value) 0.4096 chain probability
task succeed

---

start at: ((1, 2), 0) 0.280948121919 (objective value)
through policy action: E → ((1, 3), 2) 0.480205370842 (objective value) 0.8 chain probability
through policy action: W → ((1, 2), 2) 0.571854875951 (objective value) 0.64 chain probability
through policy action: W → ((1, 1), 2) 0.847764807959 (objective value) 0.512 chain probability
through policy action: S → ((2, 1), 3) 1 (objective value) 0.4096 chain probability
task succeed

---

start at: ((1, 3), 0) 0.270597074631 (objective value)
through policy action: S → ((2, 3), 0) 0.280948121919 (objective value) 0.8 chain probability
through policy action: N → ((1, 3), 2) 0.480205370842 (objective value) 0.64 chain probability
through policy action: W → ((1, 2), 2) 0.571854875951 (objective value) 0.512 chain probability
through policy action: W → ((1, 1), 2) 0.847764807959 (objective value) 0.4096 chain probability
through policy action: S → ((2, 1), 3) 1 (objective value) 0.32768 chain probability
task succeed

---

start at: ((2, 1), 0) 0.169257473733 (objective value)
through policy action: N → ((1, 1), 0) 0.31484740392 (objective value) 0.8 chain probability
through policy action: S → ((2, 1), 1) 0.362877291661 (objective value) 0.64 chain probability
through policy action: N → ((1, 1), 1) 0.583646203241 (objective value) 0.512 chain probability
through policy action: E → ((1, 2), 1) 0.692307692307 (objective value) 0.4096 chain probability
through policy action: E → ((1, 3), 3) 1 (objective value) 0.32768 chain probability
task succeed

---

start at: ((2, 2), 0) 0.260043854771 (objective value)
through policy action: N → ((1, 2), 0) 0.280948121919 (objective value) 0.8 chain probability
through policy action: E → ((1, 3), 2) 0.480205370842 (objective value) 0.64 chain probability
through policy action: W → ((1, 2), 2) 0.571854875951 (objective value) 0.512 chain probability
through policy action: W → ((1, 1), 2) 0.847764807959 (objective value) 0.4096 chain probability
through policy action: S → ((2, 1), 3) 1 (objective value) 0.32768 chain probability
task succeed

---

start at: ((2, 3), 0) 0.280948121919 (objective value)
through policy action: N → ((1, 3), 2) 0.480205370842 (objective value) 0.8 chain probability
through policy action: W → ((1, 2), 2) 0.571854875951 (objective value) 0.64 chain probability
through policy action: W → ((1, 1), 2) 0.847764807959 (objective value) 0.512 chain probability
through policy action: S → ((2, 1), 3) 1 (objective value) 0.4096 chain probability
task succeed