Group I

Nhat Nguyen

Khuong Nguyen

Dong Nguyen

Dang Nguyen (absent)

Quan Nguyen

SARSA Algorithm

Policy: agent starts is state 1 and applies operators:

e-p-s-d-w-n-e-p-s-d-w-e-w-n-e-p-w-e

Assuming $\alpha = 0.5$ and $\gamma = 0.5$, pick up/drop off reward is 13 and the move penalty is 1. State Space: {1, 2, 3, 4, 1′, 2′, 3′, 4′}

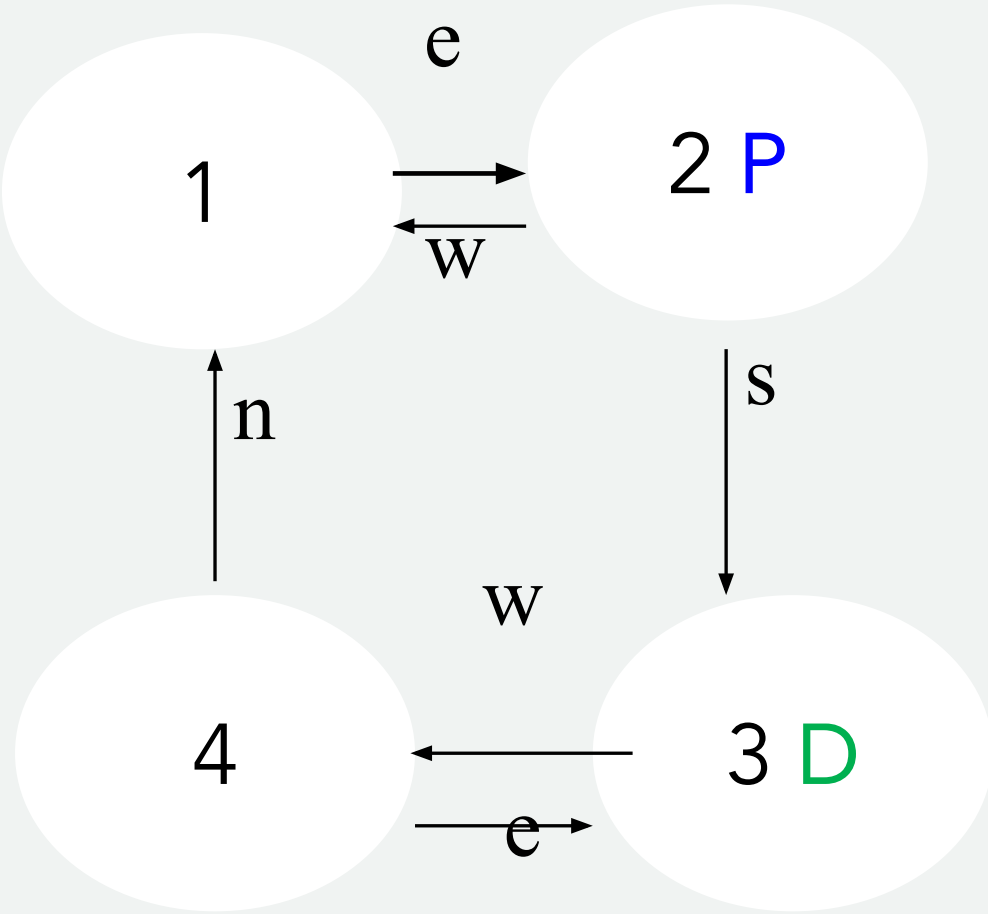
Equation:

$Q(a,s) \leftarrow Q(a,s) + \alpha[R(a,s) + \gamma\, Q(a',s') - Q(a,s)]$.

$\Rightarrow Q(a,s) \leftarrow (1- \alpha)Q(a,s)+ \alpha[R(a,s) + \gamma\, Q(a',s')]$

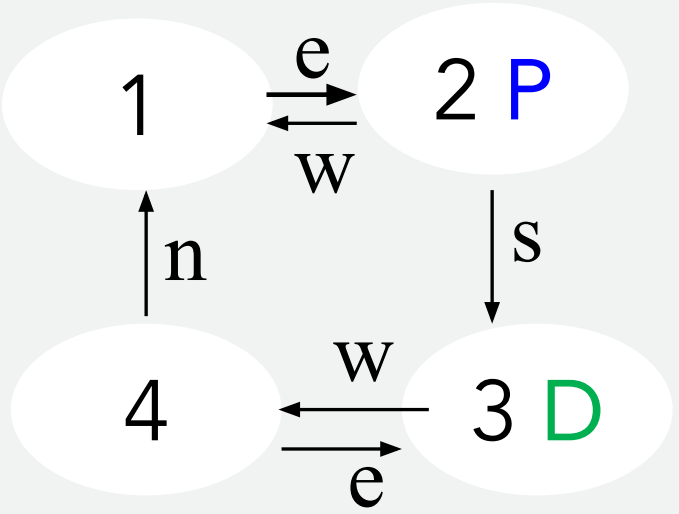
s, s′: State

a, a′: Action



Initial q-Table

	n	s	e	w	p	d
1	-	-	0	-	-	-
2	-	0	-	0	0	-
3	-	-	-	0	-	-
4	0	-	0	-	-	-
1'	-	-	0	-	-	-
2'	-	0	-	0	-	-
3'	-	-	-	0	-	0
4'	0	-	0	-	-	-



Reward Table

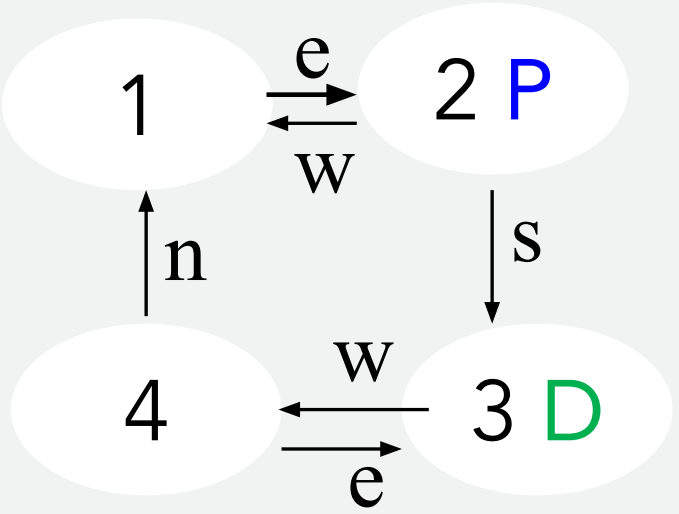
	n	s	e	w	p	d
1	-	-	-1	-	-	-
2	-	-1	-	-1	13	-
3	-	-	-	-1	-	-
4	-1	-	-1	-	-	-
1'	-	-	-1	-	-	-
2'	-	-1	-	-1	-	-
3'	-	-	-	-1	-	13
4'	-1	-	-1	-	-	-

Updated q-Table

	n	s	e	w	p	d
1	-	-	-0.5	-	-	-
2	-	0	-	0	0	-
3	-	-	-	0	-	-
4	0	-	0	-	-	-
1'	-	-	0	-	-	-
2'	-	0	-	0	-	-
3'	-	-	-	0	-	0
4'	0	-	0	-	-	-

Policy: e-p-s-d-w-n-e-p-s-d-w-e-w-n-e-p-w-e

Q(e,1) = Q(e,1)*(1 - 0.5) + 0.5*(R(e,1) + 0.5*Q(p,2)) = 0*0.5 + 0.5*(-1 + 0.5*0) = -0.5



Reward Table

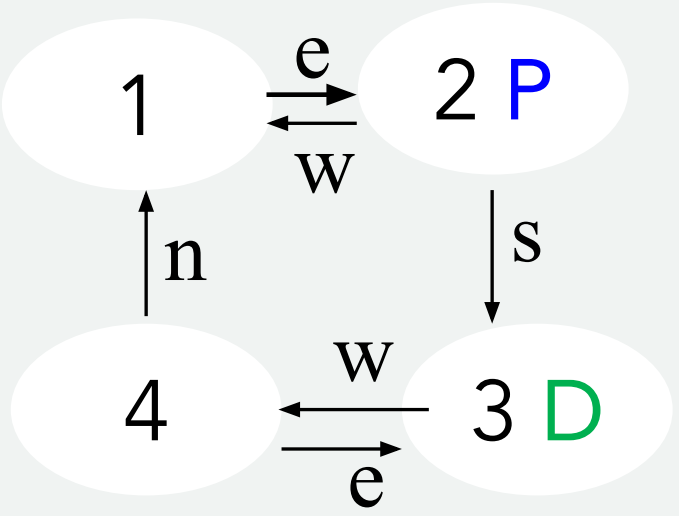
	n	s	e	w	p	d
1	-	-	-1	-	-	-
2	-	-1	-	-1	13	-
3	-	-	-	-1	-	-
4	-1	-	-1	-	-	-
1'	-	-	-1	-	-	-
2'	-	-1	-	-1	-	-
3'	-	-	-	-1	-	13
4'	-1	-	-1	-	-	-

Updated q-Table

	n	s	e	w	p	d
1	-	-	-0.5	-	-	-
2	-	0	-	0	6.5	-
3	-	-	-	0	-	-
4	0	-	0	-	-	-
1'	-	-	0	-	-	-
2'	-	0	-	0	-	-
3'	-	-	-	0	-	0
4'	0	-	0	-	-	-

Policy: e-p-s-d-w-n-e-p-s-d-w-e-w-n-e-p-w-e

Q(p,2) = Q(p,2)*(1 - 0.5) + 0.5*(R(p,2) + 0.5*Q(s,2')) = 0*0.5 + 0.5*(13 + 0.5*0) = 6.5



Reward Table

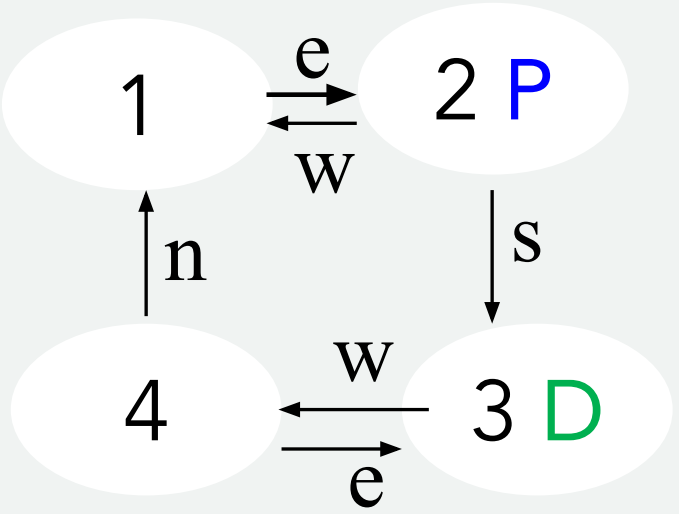
	n	s	e	w	p	d
1	-	-	-1	-	-	-
2	-	-1	-	-1	13	-
3	-	-	-	-1	-	-
4	-1	-	-1	-	-	-
1'	-	-	-1	-	-	-
2'	-	-1	-	-1	-	-
3'	-	-	-	-1	-	13
4'	-1	-	-1	-	-	-

Updated q-Table

	n	s	e	w	p	d
1	-	-	-0.5	-	-	-
2	-	0	-	0	6.5	-
3	-	-	-	0	-	-
4	0	-	0	-	-	-
1'	-	-	0	-	-	-
2'	-	-0.5	-	0	-	-
3'	-	-	-	0	-	0
4'	0	-	0	-	-	-

Policy: e-p-s-d-w-n-e-p-s-d-w-e-w-n-e-p-w-e

Q(s,2') = Q(s,2')*(1 - 0.5) + 0.5*(R(s,2') + 0.5*Q(d,3')) = 0*0.5 + 0.5*(-1 + 0.5*0) = -0.5



Reward Table

	n	s	e	w	p	d
1	-	-	-1	-	-	-
2	-	-1	-	-1	13	-
3	-	-	-	-1	-	-
4	-1	-	-1	-	-	-
1'	-	-	-1	-	-	-
2'	-	-1	-	-1	-	-
3'	-	-	-	-1	-	13
4'	-1	-	-1	-	-	-

Calculations:

$Q(e,1) = Q(e,1)*(1 - 0.5) + 0.5*(R(e,1) + 0.5*Q(p,2)) = 0*0.5 + 0.5*(-1 + 0.5*0) = -0.5$

$Q(p,2) = Q(p,2)*(1 - 0.5) + 0.5*(R(p,2) + 0.5*Q(s,2')) = 0*0.5 + 0.5*(13 + 0.5*0) = 6.5$

$Q(s,2') = Q(s,2')*(1 - 0.5) + 0.5*(R(s,2') + 0.5*Q(d,3')) = 0*0.5 + 0.5*(-1 + 0.5*0) = -0.5$

$Q(d,3') = Q(d,3')*(1 - 0.5) + 0.5*(R(d,3') + 0.5*Q(w,3)) = 0*0.5 + 0.5*(13 + 0.5*0) = 6.5$

$Q(w,3) = Q(w,3)*(1 - 0.5) + 0.5*(R(w,3) + 0.5*Q(n,4)) = 0*0.5 + 0.5*(-1 + 0.5*0) = -0.5$

$Q(n,4) = Q(n,4)*(1 - 0.5) + 0.5*(R(n,4) + 0.5*Q(e,1)) = 0*0.5 + 0.5*(-1 + 0.5*-0.5) = -0.625$

$Q(e,1) = Q(e,1)*(1 - 0.5) + 0.5*(R(e,1) + 0.5*Q(p,2)) = -0.5*0.5 + 0.5*(-1 + 0.5*6.5) = 0.875$

$Q(p,2) = Q(p,2)*(1 - 0.5) + 0.5*(R(p,2) + 0.5*Q(s,2')) = 6.5*0.5 + 0.5*(13 + 0.5*-0.5) = 9.625$

$Q(s,2') = Q(s,2')*(1 - 0.5) + 0.5*(R(s,2') + 0.5*Q(d,3')) = -0.5*0.5 + 0.5*(-1 + 0.5*6.5) = 0.875$

Continue...

Q(d,3') = Q(d,3')*(1 - 0.5) + 0.5*(R(d,3') + 0.5*Q(w,3)) = 6.5*0.5 + 0.5*(13 + 0.5*-0.5) = 9.625

Q(w,3) = Q(w,3)*(1 - 0.5) + 0.5*(R(w,3) + 0.5*Q(e,4)) = -0.5*0.5 + 0.5*(-1 + 0.5*0) = -0.75

Q(e,4) = Q(e,4)*(1 - 0.5) + 0.5*(R(e,4) + 0.5*Q(w,3)) = 0*0.5 + 0.5*(-1 + 0.5*-0.75) = -0.6875

Q(w,3) = Q(w,3)*(1 - 0.5) + 0.5*(R(w,3) + 0.5*Q(n,4)) = -0.75*0.5 + 0.5*(-1 + 0.5*-0.625) = -1.03125

Q(n,4) = Q(n,4)*(1 - 0.5) + 0.5*(R(n,4) + 0.5*Q(e,1)) = -0.625*0.5 + 0.5*(-1 + 0.5*0.875) = -0.59375

Q(e,1) = Q(e,1)*(1 - 0.5) + 0.5*(R(e,1) + 0.5*Q(p,2)) = 0.875*0.5 + 0.5*(-1 + 0.5*9.625) = 2.34375

Q(p,2) = Q(p,2)*(1 - 0.5) + 0.5*(R(p,2) + 0.5*Q(w,2')) = 9.625*0.5 + 0.5*(13 + 0.5*0) = 11.3125

Q(w,2') = Q(w,2')*(1 - 0.5) + 0.5*(R(w,2') + 0.5*Q(e,1')) = 0*0.5 + 0.5*(-1 + 0.5*0) = -0.5

Q(e,1') = Q(e,1')*(1 - 0.5) + 0.5*(R(e,1') + 0.5***0**) = 0*0.5 + 0.5*(-1 + 0.5*0) = -0.5

Final q-Table

	n	s	e	w	p	d
1	-	-	2.34375	-	-	-
2	-	0	-	0	11.3125	-
3	-	-	-	-1.03125	-	-
4	-0.59375	-	-0.6875	-	-	-
1’	-	-	-0.5	-	-	-
2’	-	0.875	-	-0.5	-	-
3’	-	-	-	0	-	9.625
4’	0	-	0	-	-	-