

✓ Discrete Unicycle Environment and Q-Learning Agent

✓ Introduction

This notebook defines a custom discrete grid-based environment for a unicycle agent and implements a Q-learning agent to learn navigation towards a goal while avoiding obstacles.

✓ Environment Class Definition

The `DiscreteUnicycleEnv` class defines a grid-based environment where the agent navigates to a goal while avoiding obstacles.

Rewards:

Reach Destination = +100

Distance to Target= -distance

Collide with obstacles = -100

number of steps = -number

End Episode:

Reach Target

Collide with obstacles

```
1 import numpy as np
2 import matplotlib.pyplot as plt
3 from tqdm import tqdm
4 from collections import deque
5 from matplotlib.patches import Rectangle, Arrow
6 from IPython.display import display, clear_output
7 import gym
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
```

```

20 self.relative_state = self.get_relative_state()
21 self.obstacles = np.zeros((grid_size, grid_size), dtype=np.int32)
22 if obstacles:
23     for obs in obstacles:
24         self.obstacles[obs] = 1
25
26 self.path = []
27 self.fig, self.ax = plt.subplots()
28 self.max_steps_per_episode = 50 # Maximum steps per episode
29
30 def reset(self):
31     self.state = self.initial_state.copy() # Reset to initial state
32
33     self.relative_state = self.get_relative_state()
34     self.path = [self.state[:2].copy()]
35     return {'state': self.state, 'obstacles': self.obstacles}
36 #####
37 def step(self, action):
38     x, y, theta = self.state
39     if action == 0: # Move right
40         x += 1
41     elif action == 1: # Move up-right
42         x += 1
43         y += 1
44     elif action == 2: # Move up
45         y += 1
46     elif action == 3: # Move up-left
47         x -= 1
48         y += 1
49     elif action == 4: # Move left
50         x -= 1
51     elif action == 5: # Move down-left
52         x -= 1
53         y -= 1
54     elif action == 6: # Move down
55         y -= 1
56     elif action == 7: # Move down-right
57         x += 1
58         y -= 1
59
60     x = np.clip(x, 0, self.grid_size - 1)
61     y = np.clip(y, 0, self.grid_size - 1)
62     self.state = np.array([x, y, theta])
63
64     self.relative_state = self.get_relative_state()
65     self.path.append(self.state[:2].copy())
66
67     done = self.is_done()
68     reward = self.get_reward()
69
70     return {'state': self.state, 'obstacles': self.obstacles}, reward, done, {}
71
72 def get_relative_state(self):
73     x, y, theta = self.state
74     goal_x, goal_y, goal_theta = self.goal
75     dx = goal_x - x
76     dy = goal_y - y
77     ex = (np.cos(theta * np.pi / 4) * dx + np.sin(theta * np.pi / 4) * dy) * 2
78     ey = (-np.sin(theta * np.pi / 4) * dx + np.cos(theta * np.pi / 4) * dy) * 2
79     ex = np rint(ex)
80     ey = np rint(ey)
81     etheta = (goal_theta - theta) % 8
82     return np.array([ex, ey, etheta], dtype=np.int16)
83 #####

```

```

84 def get_reward(self):
85     x, y, theta = self.state
86     if self.obstacles[x, y] == 1: # Collision with obstacle
87         return -100
88     distance = np.sqrt((x - self.goal[0])**2 + (y - self.goal[1])**2)
89     reward = -distance
90     if self.is_done():
91         reward += 100
92     else:
93         reward -= 1 # Penalty for each step
94     return reward
95 #####
96 def is_done(self):
97     x, y, theta = self.state
98     if self.obstacles[x, y] == 1: # Collision with obstacle
99         return True
100     return np.array_equal(self.state[:2], self.goal[:2])
101
102 def render(self, mode='human'):
103     x, y, theta = self.state
104     goal_x, goal_y, goal_theta = self.goal
105     self.ax.clear()
106     buffer = 1
107     self.ax.set_xlim(-buffer, self.max_x + buffer)
108     self.ax.set_ylim(-buffer, self.max_y + buffer)
109     self.ax.add_patch(Rectangle((-buffer, -buffer), self.max_x + 2*buffer, self.max_y + 2*buffer, fill=None, edgecolor='gray', linestyle='-', linewidth=1))
110     self.ax.plot(goal_x, goal_y, 'ro', label='Goal')
111     self.ax.add_patch(Arrow(goal_x, goal_y, np.cos(goal_theta * np.pi / 4), np.sin(goal_theta * np.pi / 4), width=0.5, color='r'))
112     self.ax.plot(self.path[0][0], self.path[0][1], 'go', label='Start')
113 #####
114     # Plot obstacles
115     for i in range(self.grid_size):
116         for j in range(self.grid_size):
117             if self.obstacles[i, j] == 1:
118                 self.ax.add_patch(Rectangle((i, j), 1, 1, color='gray'))
119
120     path = np.array(self.path)
121     self.ax.plot(path[:, 0], path[:, 1], 'k--', label='Path')
122     width, height = self.max_x / 20, self.max_x / 30
123     robot = Rectangle((x - 0.5 * width, y - 0.5 * height), width, height, angle=np.degrees(theta * np.pi / 4), edgecolor='b', facecolor='b')
124     self.ax.add_patch(robot)
125     self.ax.add_patch(Arrow(x, y, np.cos(theta * np.pi / 4), np.sin(theta * np.pi / 4), width=0.5, color='b'))
126     self.ax.set_aspect('equal', adjustable='box')
127     self.ax.legend()
128     self.ax.grid(True)
129     clear_output(wait=True)
130     display(self.fig)
131     plt.pause(0.001)
132
133
134 def close(self):
135     plt.close(self.fig)
136 #####
137 # Define obstacles
138 obstacles = [(0,1),(1,1),(2,1),(3,1),(2,7),(4,3),(5,6),(7,3),(8,5),(6,3),(8,5),(7,6),(7,5),(4,7),(7,4),(1,7),(1,6),(1,5)]
139
140 # Create environment instance
141 env = DiscreteUnicycleEnv(obstacles=obstacles)

```



Show hidden output

✓ Q-Learning Agent Class Definition

The QLearningAgent class defines an agent that uses Q-learning to learn to navigate the DiscreteUnicycleEnv environment.

```
1 class QLearningAgent:
2     def __init__(self, state_size, action_size, learning_rate=0.1, discount_factor=0.95, epsilon=1.0, epsilon_decay=0.995, epsilon_min=0.01):
3         self.state_size = state_size
4         self.action_size = action_size
5         self.learning_rate = learning_rate
6         self.discount_factor = discount_factor
7         self.epsilon = epsilon
8         self.epsilon_decay = epsilon_decay
9         self.epsilon_min = epsilon_min
10        self.q_table = np.zeros(state_size + (action_size,))
11
12    def choose_action(self, state):
13        if np.random.rand() <= self.epsilon:
14            return np.random.choice(self.action_size)
15        return np.argmax(self.q_table[state['state'][0], state['state'][1], state['state'][2]])
16
17    def learn(self, state, action, reward, next_state, done):
18        q_update = reward
19        if not done:
20            q_update += self.discount_factor * np.max(self.q_table[next_state['state'][0], next_state['state'][1], next_state['state'][2]])
21        self.q_table[state['state'][0], state['state'][1], state['state'][2], action] += self.learning_rate * (q_update - self.q_table[state['state'][0], state['state'][1], state['state'][2], action])
22        if self.epsilon > self.epsilon_min:
23            self.epsilon *= self.epsilon_decay
24
25    # Define state and action sizes
26    state_size = (env.grid_size, env.grid_size, 8)
27    action_size = env.action_space.n
28
29    # Create Q-learning agent instance
30    agent = QLearningAgent(state_size, action_size)
31
```

✓ Training the Q-Learning Agent

Train the Q-learning agent on the DiscreteUnicycleEnv environment for a specified number of episodes.

```
1 episodes = 5000
2 max_steps = 50
3
4 average_rewards = []
5
6 for e in tqdm(range(episodes), desc="Training Episodes"):
7     state = env.reset()
8     total_reward = 0
9     for _ in range(max_steps):
10        action = agent.choose_action(state)
11        next_state, reward, done, _ = env.step(action)
12        agent.learn(state, action, reward, next_state, done)
13        state = next_state
14        total_reward += reward
15        if done:
16            break
17    average_rewards.append(total_reward / max_steps)
18
```

Training Episodes: 100%|██████████| 5000/5000 [00:08<00:00, 566.08it/s]

✓ Evaluating the Trained Agent

Run the trained Q-learning agent in the environment to visualize its performance.

```
1 # Render environment for visual display
2 state = env.reset()
3 done = False
4 while not done:
5     action = agent.choose_action(state)
6     next_state, reward, done, _ = env.step(action)
7     env.render()
8     state = next_state
9 env.close()
```

