Colab 連結:

https://colab.research.google.com/drive/1N-AT8PmveTvGcjchtrjqIJ2-DoG6j57X?usp=sharing

Q-table

Q-table:

(<built-in function all>,	left	right	up	down
0	-38.665820	-22.265811	-38.875258	-22.256859
1	-22.272731	-22.272011	-36.209018	-36.238582
2	-22.278917	-22.278749	-37.968134	-37.585777
3	-22.285670	-37.301270	-37.788665	-22.285625
4	0.000000	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.000000
6	-26.249561	-23.969910	-25.669642	-22.322952
7	0.000000	0.000000	0.000000	0.000000
8	-24.731756	-24.201207	-24.970259	-22.342267
9	0.000000	0.000000	0.000000	0.000000
10	-25.350170	-22.297517	-23.513977	-24.856502
11	-22.295116	-22.280806	-45.559816	-22.295065
12	-22.300196	-18.146104	-52.760838	-22.300245
13	-22.230202	-20.437097	-57.786308	-22.343084
14	-22.494627	-20.837470	-55.853596	-55.303626
15	-22.785723	-21.258699	-59.210366	-22.785891
16	-22.995558	-21.414556	-59.229562	-22.995276
17	-23.147344	-21.511478	-59.169969	-23.148555
18	-23.365011	-22.647781	-58.205520	-57.042117
19	-23.807711	-23.648848	-61.478436	-23.816109
20	-24.332915	-50.785292	-50.588746	-24.333482
21	-36.790772	-36.914863	-22.266959	-22.257822
22	0.000000	0.000000	0.000000	0.000000
23	0.000000	0.000000	0.000000	0.000000
24	-24.180412	-24.529398	-22.292241	-26.755278
25	0.000000	0.000000	0.000000	0.000000
26	-23.715359	-22.323068	-23.814221	-24.824828
27	-22.313471	-22.313491	-22.313342	-22.314042
28	-22.323493	-22.323134	-36.948009	-38.564152
29	-22.333092	-37.727309	-22.332564	-36.568266
30	0.000000	0.000000	0.000000	0.000000
31	0.000000	0.000000	0.000000	0.000000
32	-38.104021	-22.295689	-22.294438	-37.320965
33	-22.292090	-22.292619	-22.123128	-22.293480
34	-22.305425	-35.630270	-22.301019	-36.175469
35	0.000000	0.000000	0.000000	0.000000
36	-38.918600	-22.823593	-22.822667	-37.947645
37	-22.944799	-22.945073	-22.945100	-48.396500
38	-23.043492	-38.417963	-23.043249	-37.253706
39	0.000000	0.000000	0.000000	0.000000
40	-42.741499	-24.001435	-24.001315	-40.538779

41	-24.160375	-50.898090	-24.160549	-24.160783
42	-36.916269	-36.475002	-22.267828	-22.258309
43	0.000000	0.000000	0.000000	0.000000
44	0.000000	0.000000	0.000000	0.000000
45	0.000000	0.000000	0.000000	0.000000
46	-24.358408	-22.675008	-23.832044	-22.305475
47	0.000000	0.000000	0.000000	0.000000
48	-37.284443	-37.926322	-22.303900	-22.305516
49	0.000000	0.000000	0.000000	0.000000
50	0.000000	0.000000	0.000000	0.000000
51	0.000000	0.000000	0.000000	0.000000
52	-24.304966	-25.262812	-24.278704	-22.353293
53	0.000000	0.000000	0.000000	0.000000
54	-37.879853	-36.894878	-22.278233	-22.296106
55	0.000000	0.000000	0.000000	0.000000
56	-23.435725	-23.961170	-24.768182	-22.312146
57	0.000000	0.000000	0.000000	0.000000
58	0.000000	0.000000	0.000000	0.000000
59	0.000000	0.000000	0.000000	0.000000
60	-29.464628	-27.620822	-27.817358	-24.263476
61	0.000000	0.000000	0.000000	0.000000
62	-42.119758	-40.896056	-24.168336	-24.167757
63	-46.626351	-22.264899	-22.268421	-22.252675
64	-22.271517	-24.738665	-25.878628	-24.065056
65	0.000000	0.000000	0.000000	0.000000
66	-37.511583	-22.285622	-36.803106	-22.296628
67	-22.295907	-22.286312	-22.293815	-45.856467
68	-22.296492	-22.287203	-45.714732	-22.294751
69	-22.297423	-22.286949	-22.295423	-46.558414
70	-22.297067	-37.523093	-37.382449	-22.287812
71	0.000000	0.000000	0.000000	0.000000
72	-37.435099	-22.330907	-36.653931	-22.331338
73	-22.342432	-37.760494	-22.342069	-38.353995
74	0.000000	0.000000	0.000000	0.000000
75	-47.068618	-22.293607	-22.285796	-22.298226
76	-22.300409	-22.300103	-35.902969	-36.592421
77	-22.307080	-22.306369	-22.305742	-22.306115
78	-22.309686	-22.309218	-45.606470	-22.309002
79	-22.315633	-24.590683	-24.491487	-24.184849
80	0.000000	0.000000	0.000000	0.000000

81	-41.776690	-24.224181	-24.224135	-40.815245
82	-24.184043	-24.183579	-50.764856	-24.184704
83	-24.175953	-40.144661	-24.175360	-40.259137
84	-36.425701	-37.483788	-22.268288	-22.069020
85	0.000000	0.000000	0.000000	0.000000
86	-37.327496	-22.291978	-35.564774	-22.297857
87	-22.296419	-45.137689	-22.286414	-22.293929
88	0.000000	0.000000	0.000000	0.000000
89	-25.106766	-25.804185	-22.308299	-23.955683
90	0.000000	0.000000	0.000000	0.000000
91	-46.461402	-22.296906	-22.298029	-22.291401
92	-22.308758	-22.308441	-37.139925	-37.745320
93	-22.320185	-37.697948	-22.319701	-36.170548
94	0.000000	0.000000	0.000000	0.000000
95	-38.214899	-22.299682	-36.953776	-22.299397
96	-22.295261	-46.626331	-22.287876	-22.296025
97	0.000000	0.000000	0.000000	0.000000
98	-35.302310	-22.308820	-22.309698	-37.336790
99	-22.309293	-37.479349	-22.308715	-36.505780
100	0.000000	0.000000	0.000000	0.000000
101	0.000000	0.000000	0.000000	0.000000
102	0.000000	0.000000	0.000000	0.000000
103	-40.697814	-41.237957	-24.192161	-24.192824
104	0.000000	0.000000	0.000000	0.000000
105	-52.582516	-17.838921	-22.274379	-22.271723
106	-22.295417	-22.287837	-47.785259	-45.519315
107	-22.297600	-22.296976	-22.288940	-45.491382
108	-22.299650	-36.522020	-22.300225	-38.252288
109	0.000000	0.000000	0.000000	0.000000
110	0.000000	0.000000	0.000000	0.000000
111	0.000000	0.000000	0.000000	0.000000
112	-37.459504	-36.460209	-22.301501	-22.289370
113	0.000000	0.000000	0.000000	0.000000
114	0.000000	0.000000	0.000000	0.000000
115	-24.075890	-22.308024	-24.846910	-23.565117
116	-22.298559	-22.300186	-22.298816	-22.299315
117	-22.297095	-22.296748	-22.288395	-22.299109
118	-22.299562	-38.051627	-37.110243	-22.300526
119	0.000000	0.000000	0.000000	0.000000
120	0.000000	0.000000	0.000000	0.000000

121	-26.398013	-24.314046	-27.734894	-27.275401
122	-24.277086	-24.276363	-42.373509	-42.302646
123	-24.238683	-24.239129	-51.277370	-24.239101
124	-24.200851	-24.201445	-24.200390	-48.957803
125	-24.208854	-40.873284	-41.930469	-24.209131
126	-37.305647	-36.604150	-22.260864	-22.270714
127	0.000000	0.000000	0.000000	0.000000
128	0.000000	0.000000	0.000000	0.000000
129	0.000000	0.000000	0.000000	0.000000
130	-24.080923	-22.336493	-24.438576	-25.320536
131	-22.325512	-37.187343	-37.213378	-22.325646
132	0.000000	0.000000	0.000000	0.000000
133	-37.311329	-36.142580	-22.299499	-22.291171
134	0.000000	0.000000	0.000000	0.000000
135	-24.755039	-23.843706	-23.764982	-22.359548
136	0.000000	0.000000	0.000000	0.000000
137	-36.324428	-22.299114	-22.299083	-37.076615
138	-22.298555	-22.298720	-22.291693	-22.302395
139	-22.300262	-22.300117	-22.299407	-22.300432
140	-22.303809	-37.643575	-36.707513	-22.304704
141	0.000000	0.000000	0.000000	0.000000
142	0.000000	0.000000	0.000000	0.000000
143	0.000000	0.000000	0.000000	0.000000
144	-40.734088	-42.186463	-24.276729	-24.276220
145	0.000000	0.000000	0.000000	0.000000
146	-40.990935	-42.005536	-24.216562	-24.217068
147	-46.256732	-22.272187	-22.272609	-22.271757
148	-22.273939	-22.273206	-36.279784	-36.589938
149	-22.275094	-22.275005	-46.222610	-22.274504
150	-22.275466	-38.347456	-37.505355	-22.276187
151	0.000000	0.000000	0.000000	0.000000
152	-37.190640	-36.739983	-22.314335	-22.314640
153	0.000000	0.000000	0.000000	0.000000
154	-35.692270	-37.006380	-22.301345	-22.292033
155	0.000000	0.000000	0.000000	0.000000
156	-36.420653	-35.844619	-22.343270	-22.343927
157	0.000000	0.000000	0.000000	0.000000
158	0.000000	0.000000	0.000000	0.000000
159	-46.102683	-22.301037	-22.295657	-22.301806
160	-22.303938	-22.302951	-22.303159	-46.491163

161	-22.303178	-22.302505	-22.302232	-22.303056
162	-22.306359	-22.306503	-36.670671	-38.112371
163	-22.308020	-36.673566	-38.042861	-22.306508
164	0.000000	0.000000	0.000000	0.000000
165	-28.541851	-27.919445	-24.313248	-28.904553
166	0.000000	0.000000	0.000000	0.000000
167	-42.207399	-40.378051	-24.224473	-24.224806
168	-36.285207	-37.491042	-22.273475	-22.272843
169	0.000000	0.000000	0.000000	0.000000
170	-36.133340	-22.276539	-22.275758	-36.494459
171	-22.276397	-46.687391	-22.276539	-22.276436
172	0.000000	0.000000	0.000000	0.000000
173	-46.428388	-22.307318	-22.303293	-22.296566
174	-22.297127	-22.302229	-37.086404	-37.025634
175	-22.292266	-35.685780	-22.302314	-37.075121
176	0.000000	0.000000	0.000000	0.000000
177	-35.568349	-36.938744	-22.327156	-22.327858
178	0.000000	0.000000	0.000000	0.000000
179	-36.670007	-22.295979	-37.997029	-22.305518
180	-22.302469	-46.421969	-22.294435	-22.301928
181	0.000000	0.000000	0.000000	0.000000
182	-36.174514	-36.891333	-22.307364	-22.307754
183	0.000000	0.000000	0.000000	0.000000
184	-46.009868	-22.308086	-22.307988	-22.307664
185	-22.319406	-23.200078	-24.426382	-24.503118
186	0.000000	0.000000	0.000000	0.000000
187	0.000000	0.000000	0.000000	0.000000
188	-40.537482	-42.072370	-24.232210	-24.232298
189	-37.242467	-37.539351	-22.274643	-22.273845
190	0.000000	0.000000	0.000000	0.000000
191	0.000000	0.000000	0.000000	0.000000
192	-36.736252	-37.159288	-22.278021	-22.277677
193	0.000000	0.000000	0.000000	0.000000
194	-38.611595	-37.296486	-22.306535	-22.298817
195	0.000000	0.000000	0.000000	0.000000
196	0.000000	0.000000	0.000000	0.000000
197	-36.643759	-22.316482	-36.246418	-22.316816
198	-22.311490	-22.310795	-22.310940	-22.311667
199	-22.310221	-22.309947	-45.381546	-22.310487
200	-22.306451	-22.303953	-22.295162	-22.305178
201	-22.304667	-22.304850	-22.304006	-22.304701
202	-22.305644	-22.306043	-46.458131	-22.305461
203	-22.305481	-22.305017	-22.304835	-22.304386
204	-22.306150	-22.306729	-46.573354	-22.306726
205	-22.308153	-36.796165	-22.308961	-37.749592
206	0.000000	0.000000	0.000000	0.000000
207	-41.441186	-24.314949	-40.925550	-24.314990
208	-24.277017	-24.277604	-41.965127	-41.764022
209	-24.239139	-41.414689	-24.239676	-1093.761658
210	-25.311772	-25.131940	-22.275392	-26.145506
211	0.000000	0.000000	0.000000	0.000000
212	-24.687480	-22.280191	-25.799038	-25.081075
213	-22.278653	-37.191258	-22.279426	-38.087005
214	0.000000	0.000000	0.000000	0.000000
215	-37.482508	-22.299693	-22.308933	-37.081050
216	-22.309785	-22.304007	-35.915530	-37.971376
217	-22.314163	-22.302951	-35.387121	-38.255788
218	-22.313023	-22.307576	-22.312160	-46.713295
219	-22.314208	-22.309551	-22.311176	-46.111466
220	-22.314329	-22.303382	-22.308789	-45.638967
221	-22.307131	-22.305956	-22.300691	-46.294791
222	-22.307203	-22.307192	-22.307763	-45.975024
223	-22.306509	-22.306001	-22.306719	-45.689247
224	-22.305757	-22.305618	-22.305954	-47.697413
225	-22.308932	-37.930549	-22.308520	-36.566703
226	0.000000	0.000000	0.000000	0.000000
227	-27.682259	-24.388731	-26.776372	-26.899492
228	-24.351978	-42.229931	-24.352383	-41.629022
229	0.000000	0.000000	0.000000	0.000000
230	0.000000	0.000000	0.000000	0.000000)

步數最少及寶藏最多的截圖

```
score為5的情況:第 749 步數 623
['Episode 749: total_steps=623 score =5']
```

步數最少 623 步有五個寶藏

Reward 設定

```
N_STATES_x = 21
N_STATES_y = 11
ACTIONS = ["left", "right", "up", "down"]
GOAL = 230
EPSILON = 0.9
ALPHA = 0.001
GAMMA = 0.95
MAX_EPISODES = 1000
FRESH_TIME = 0.1

WALL = [4, 5, 7, 9, 22, 23, 25, 30, 31, 35, 39, 43, 45, 47, 49, 50, 51, 53, 55, 57, 58, 59, 61, 65, 71, 74, 80, 85, 88,
        90, 94, 97, 100, 101, 102, 104, 109, 110, 111, 113, 114, 119, 120, 127, 128, 129, 132, 134, 136, 141, 142, 143,
        145, 151, 153, 155, 157, 158, 164, 166, 169, 172, 176, 178, 181, 183, 186, 187, 190, 191, 193, 195, 196, 206, 211,
        214, 226, 229]
CHEST = [6, 79, 170, 212, 227]

RIGHT_EXTRA = [105, 12, 13, 14, 15, 16, 17, 18, 19]

MAX_STEP = 100000
SCORE_RATIO = 0

HIGHEST_EPISODE = 0
LOWEST_STEP = 100000

HIGHEST_EPISODE_5 = 0
LOWEST_STEP_5 = 100000


SCORE = 0
```

ALPHA 設小一點降低學習率

四個動作分別都有走到 terminal、牆壁、寶藏、現在記錄的路徑、單純移動的 reward，不一樣的是我觀察到 goal 的左邊是牆壁，所以我將由上往下走到終點的動作相比往右走的 reward 設的還要高一點，目的是可以更加獎勵往下走的 reward，但兩個都設負數是希望不要全力都放在走到終點上。其他如撞牆、走到已經走過的位置都設負數，希望他能學習不要做這些行為。另外就是遇到獎勵就獎勵 300，畢竟五個寶藏都找到分數比較高。

```
if S_ in CHEST:
        CHEST.remove(S_)#移除寶藏

else:
    if S_ != 'terminal':
            q_target = R + GAMMA * q_table.iloc[S_, :].max()
    else:
            if(SCORE<5):
                R-=1500
            q_target = R
            is_terminated = True
    q_table.loc[S, A] += ALPHA * (q_target - q_predict)
    S = S_
    update_env(S, episode, step_counter + 1, path)
    step_counter += 1
```

還有沒全部找到寶藏就扣 reward

```
def get_env_feedback(S, A, path):
    global CHEST, SCORE, WALL, GOAL

    if A == "right":
        if S == GOAL - 1:
            S_ = "terminal"
            R = -1000

        elif (S % N_STATES_x == N_STATES_x - 1) or (S+1 in WALL):
            S_ = S
            R = -50
        elif S+1 in CHEST:
            S_ = S+1
            R = 300
            SCORE += 1
        elif S+1 in path:
            S_ = S+1
            R = -2
        else:
            if S in RIGHT_EXTRA:
                S_ = S + 1
                R = 50
            else:
                S_ = S + 1
                R = -1
```

```
    elif A == "left":
        if (S % N_STATES_x == 0) or (S - 1 in WALL):
            S_ = S
            R = -50
        elif S - 1 in CHEST:
            S_ = S -1
            R = 300
            SCORE += 1
        elif S - 1 in path:
            S_ = S-1
            R = -2
        else:
            S_ =S -1
            R = -1
```

```
elif A == "up":
        if (S // N_STATES_x == 0) or (S - N_STATES_x in WALL):
                S_ = S
                R = -50
        elif S - N_STATES_x in CHEST:
                S_ = S - N_STATES_x
                R = 300
                SCORE += 1
        elif S - N_STATES_x in path:
                S_ = S - N_STATES_x
                R = -2
        else:
                S_ = S - N_STATES_x
                R = -1
```

```
elif A == "down":
        if S == GOAL - N_STATES_x:
                S_ = "terminal"
                R = -500
        elif (S // N_STATES_x == N_STATES_y - 1) or (S + N_STATES_x in WALL):
                S_ = S
                R = -50
        elif S + N_STATES_x in CHEST:
                S_ = S + N_STATES_x
                R = 300
                SCORE += 1
        elif S + N_STATES_x in path:#往回走
                S_ = S + N_STATES_x
                R = -2
        else:
                S_ = S + N_STATES_x
                R = -1


return S_, R
```

心得

這次是最後一次作業，雖是 rl 常見的走迷宮題目，但我也花了一段時間才搞懂它實行的原理與實作，翻了許多的文件、網站發現能用的建議很少，也可能是查找的時間不夠，好吧其實是太晚開始做。後來決定先從迷宮本身下手，先寫再說，透過加上各種判斷，實現動作的決定。從一開始只能 1000~1200 再到 800 多再到 600 多，感動感動，但還是沒辦法到 300 步以內。跟同學討論反而收穫不少建議，如:terminal reward 設負數能更要求找到寶藏等。後再加上找到五個寶藏 R 增加，反之減少、往回走的判斷等。雖然一開始步數震盪很嚴重，但加上鼓勵找獎勵的機制後就好很多，這次的遺憾就是沒有早點開始做。