1(a)

由於output太多，這邊只截最前面跟最後面的iteration當作代表，中間的部分大同小異

一張含有 文字, 功能表, 螢幕擷取畫面, 字型 的圖片

自動產生的描述一張含有 服裝, 布, 樣式, 灰色 的圖片

自動產生的描述

(其中’X’代表足跡、’#’代表障礙物)

一張含有 文字, 螢幕擷取畫面, 字型, 功能表 的圖片

自動產生的描述一張含有 樣式, 螢幕擷取畫面, 對稱, 設計 的圖片

自動產生的描述

1(b)

一張含有 文字, 螢幕擷取畫面, 字型, 設計 的圖片

自動產生的描述

一張含有 螢幕擷取畫面, 文字, 樣式, 布 的圖片

自動產生的描述一張含有 螢幕擷取畫面, 樣式, 對稱, 設計 的圖片

自動產生的描述

1(c)

一張含有 文字, 螢幕擷取畫面, 字型, 功能表 的圖片

自動產生的描述

2(a) s1

2(b)

一張含有 文字, 字型, 螢幕擷取畫面, 設計 的圖片

自動產生的描述

3(a)

|  |  |
| --- | --- |
| State | action |
| S0 | A0 |
| S1 | A2 |
| S2 | A1 |

3(b)

|  |  |
| --- | --- |
| State | action |
| S0 | A2 |
| S1 | A2 |
| S2 | A1 |

3(c)

Gamma=0.8, Q-value:

一張含有 文字, 字型, 螢幕擷取畫面, 行 的圖片

自動產生的描述

Gamma=0.95, Q-value:

一張含有 文字, 字型, 螢幕擷取畫面, 印刷術 的圖片

自動產生的描述

|  |  |  |
| --- | --- | --- |
| State | Action(Gamma=0.8) | Action(Gamma=0.95) |
| S0 | A0 | A0 |
| S1 | A0 | A2 |
| S2 | A1 | A1 |

S1的optimal policy從a0變為a2

因為gamma增加 => 留得住的reward也變多，因此gamma=0.8時，在s1原地打轉會是最佳選擇；當gamma=0.95時，s1選擇a2走到s2被扣50，s2選擇a1走到s0得到40 reward，s0選擇a0回到原地得到10 reward，留住的reward會比在s1原地打轉更多