Artificial Intelligence

Homework 4

Reinforcement learning

姓名: 劉育辰

學號:110303585

系級:機械4C

1. State value iteration Vk\*(s)

Code:

一張含有 文字, 螢幕擷取畫面, 軟體, 字型 的圖片

AI 產生的內容可能不正確。

這裡V[]是個table會記錄每個格子的predict reward，這段會先對V[]做初始化，再來遍歷每個格子做算v-value的計算。

一張含有 文字, 螢幕擷取畫面, 軟體, 陳列 的圖片

AI 產生的內容可能不正確。

這裡是跟ppt上的計算方法一樣，先得到可能會到達的方向，再根據機率計算V-value，拿後只存取最大reward的動作，最後回傳V[]和policy table。

1. Q-value iteration Qk\*(s,a)

Code:

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

AI 產生的內容可能不正確。

因為Q-TABLE會上下左右的reward都存取，所以特別設計一個結構，紀錄每個格子的4個動作的reward和位置。

一張含有 文字, 螢幕擷取畫面, 字型, 軟體 的圖片

AI 產生的內容可能不正確。

作法上和剛剛的v-value的過程差不多，但多了q[action] = value來記錄每個action的value，這樣才畫得出和ppt上一樣的圖。

(3) Policy iteration

Code:

一張含有 文字, 螢幕擷取畫面, 字型, 軟體 的圖片

AI 產生的內容可能不正確。

這段會隨機選定policy，並對V[]做初始化。

一張含有 文字, 螢幕擷取畫面, 軟體, 陳列 的圖片

AI 產生的內容可能不正確。

這段會根據剛剛髓機給定的policy做v-value的計算。

一張含有 文字, 螢幕擷取畫面, 軟體, 陳列 的圖片

AI 產生的內容可能不正確。

這段才會開始做策略上的改進，根據每個格子的v-value來做策略的變換直到收斂。



My Calculate:

一張含有 文字, 白板, 筆跡, 設計 的圖片

AI 產生的內容可能不正確。

一張含有 文字, 筆跡, 紙張, 墨水 的圖片

AI 產生的內容可能不正確。

一張含有 文字, 筆跡, 白板, 文件 的圖片

AI 產生的內容可能不正確。

MATLAB:

一張含有 文字, 螢幕擷取畫面, 軟體, 陳列 的圖片

AI 產生的內容可能不正確。

從我的計算結果和MATLAB的結果比較可以知道，我的計算結果和V-value和Q-value相近，但是MATLAB的結果確有大於1的數值出現，可能是MATHLAB的計算方式更考慮多步累積reward得到的結果。



Result:

一張含有 文字, 螢幕擷取畫面, 正方形, Rectangle 的圖片

AI 產生的內容可能不正確。

一張含有 文字, 螢幕擷取畫面, 正方形, Rectangle 的圖片

AI 產生的內容可能不正確。

一張含有 文字, 螢幕擷取畫面, 正方形, Rectangle 的圖片

AI 產生的內容可能不正確。

Discussion:

從上面三個結果的圖可看出，V-value和Q-value的policy一樣，差在Q-value的table會多了其他三個action的值，但前面這兩個跟Policy iteration的圖有些差別，在(0,2)這格，Policy iteration認為往右比較好，再來各格子的值也和前兩個的不一樣，我用程式跑得到的是經過5次遞迴就收斂了，但是前兩個方法要經過12次左右才收斂。

Conclusion:

結論是我用code跑Q-VALUE和V-VALUE結果更接近ppt上的結果，用手算的數值也能證明我的code跑的是對的，可是透過Policy iteration卻得到不一樣的reward數值，但policy卻是是一樣的，如果要找出最佳的action，也是直線往上走，再往右走的終點。討論最後policy的決策化，最佳的入線有達到要求，所以三個方法都可行。