



摘要

在工作中的績效考核及工安表現一直是企業對員工的重要評量指標之一，但考核表中呈現出的數字或許無法讓員工有具象化的感受，因此我們希望透過深度強化學習訓練 NPC，將這些數字具體化呈現，讓員工可以看到他們在工作中的樣子。我們希望透過 NPC 讓員工對於這些數字更有感覺，知道自己在工作時的表現好壞，從而激勵他們主動追求更好的安全及紀律的紀錄數據，改善自己的缺點，達到自主管理與控制的目標，在工作中表現得更好。

第一階段實驗：訓練模型平衡

在這一階段實驗，我們先將 Walker 的最大速度設為 0，再修改原先設定的 reward，這次實驗中選擇將 match speed reward 刪除，只留下 look at target reward，並將 touch target reward 的獎勵值設為 0，希望可以藉此讓 Walker 不去觸碰目標且不移動。在訓練了 800 萬步後發現 Walker 可以很順利地站立。從下圖可以看出在 400 萬到 800 萬步間獲得的 Reward 有明顯成長，約在 700 多萬步時所累積到的 Reward 值最高。最後訓練出來的結果，Walker 是可以站著的，但手的部分仍然會有抽動的現象，不是呈現「立正」完全不會動的姿勢，雖然樣子不如動畫一樣可以與真人非常相似，但也有達成我們希望的結果。



可以看出大概在訓練至400萬步時reward明顯增加

討論與結論

我們將目前第一階段實驗中訓練出來的 Walker 與 Python 連結，觀察 Walker 做出的動作有什麼變化，再逐步定義出優等、普通、劣等的評分等級的動作。雖然我們目前做出的成果還尚在初期，Walker 調整出來的動作看起來還是很不自然，只能大致呈現出我們想表達的優劣等級，但我們希望未來能改善這些問題並做出一個動作流暢，且具有智能行為的 NPC，並調整出更加精細的動作，定義出更多等級，使我們的研究可以更加完整，並讓呈現出的結果更加生動有趣，讓員工在看到這些 Agent 的動作後可以激勵自己追求更好的評分等級。

研究方法



訓練場景

Action

共39個
hips、handL、handR不能旋轉

$$a = (\text{rotate}_x, \text{rotate}_y, \text{rotate}_z, \text{joint strength})$$

Reward

Reward 名稱	獲得 Reward 條件	Reward 值
look at target	頭部面向目標且兩者越靠近時	+0.01
match speed	人物前進方向與面向目標物的方向夾角越小時	+0.01
touch target	碰觸到目標時	+0.01

State

共243個

$$s = (bp_1 \sim bp_{16}, bp_{touching\ ground}, \Delta_x, \Delta_z, \theta)$$

第二階段實驗：訓練績效表現模型

我們在第二階段實驗中將做出 3 個評分等級的 Agent 動作，員工的表現控制他們的 Agent 會做出什麼動作。表現優等會走到目標，表現普通站在原地，表現劣等的站不起來。我們將最原始的 Walker 定義為優等的動作，第一階段實驗的結果定義為普通等級的動作，劣等的動作則利用第一階段訓練出的神經網路連結與 Python 連結。由於 Python 與 Unity 機率抽樣方式不一樣，因此我們在連結 Python 後 Agent 的表現反而變差，所以我們將其定義為劣等的動作。雖然 Agent 的動作不自然，但本階段實驗的目的是想做出各種等級的動作，讓員工知道他們的工作表現與其他表現的差異。

等級	優等	普通	劣等
動作	碰到目標	站在原地	站不起來
Agent	無修改的 Walker	第一階段實驗結果	連結 Python 後結果