**使用 Seq2Seq LSTM 與 Seq2Seq VAE 生成「未來學習行為序列」數據**

## **模型比較（MSE、Best-of-N、Coverage、Diversity）**

| 指標 | Seq2Seq LSTM | Seq2Seq VAE | 指標意義 |
| --- | --- | --- | --- |
| MSE | 以單一路徑預測未來兩週 clicks，計算整體均方誤差 | 使用多重潛在變數取樣生成 N 條未來序列，取最小 MSE（Best-of-N） | MSE 反映預測精確度；數值越小代表越準確 |
| Best-of-N MSE | — | 透過 20 次隨機採樣取最優樣本，代表生成潛力的上限 | 衡量模型在多樣生成中的最佳表現 |
| Diversity | — | 平均每筆樣本的標準差（在 20 條生成樣本間） | 反映模型生成的多樣化程度 |
| Coverage | — | 測試集中有多少比例樣本至少被某一條生成樣本覆蓋在誤差容忍範圍內 | 衡量模型是否能「涵蓋」真實分布的多樣可能性 |

## 

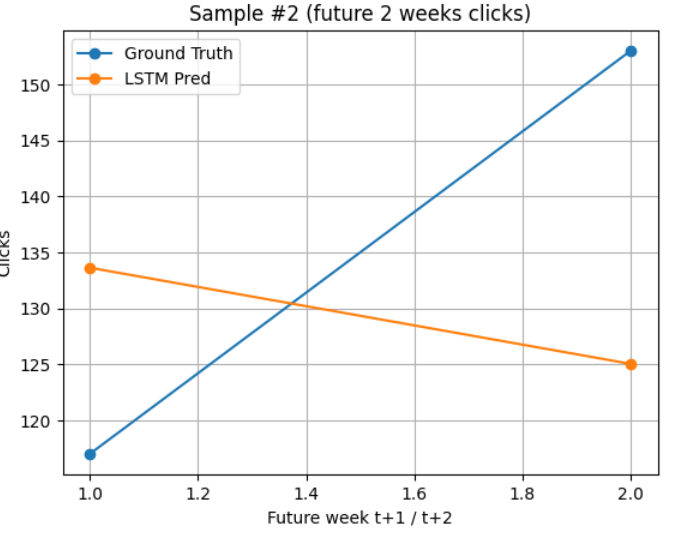
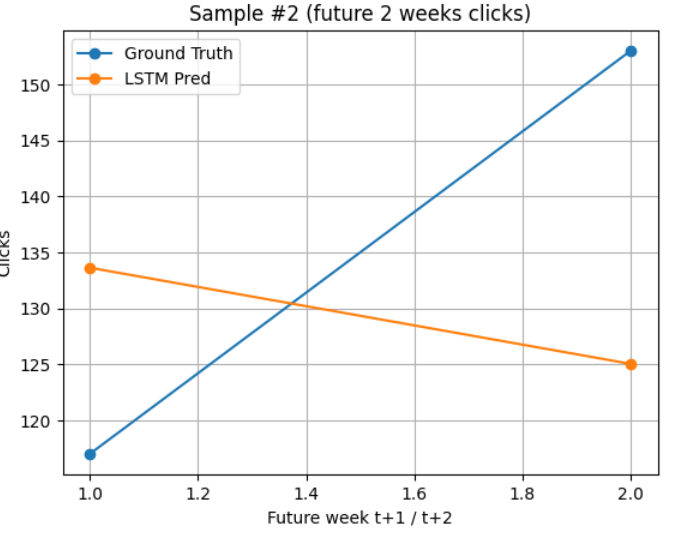
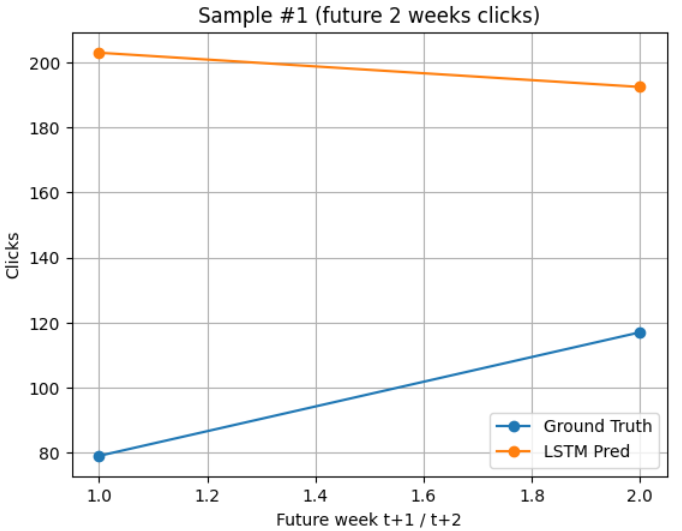
| 指標 | Seq2Seq LSTM | Seq2Seq VAE (N=20) | 說明 |
| --- | --- | --- | --- |
| MSE | 4727.9459 | 4700.0161 | VAE 的誤差略低，顯示其具潛在生成優勢 |
| Best-of-N MSE | — | 4700.0161 | 取 20 次生成樣本的最小 MSE，反映模型生成潛力上限 |
| Diversity (std) | — | 0.2362 | 生成樣本之間的標準差，代表多樣性程度 |
| Coverage (比例) | — | 0.5173 | 代表約 51.7% 測試樣本被 VAE 生成結果覆蓋在合理範圍內 |

## **整體而言，VAE 模型在準確度上雖僅略優於 LSTM（MSE 約下降 0.6%），但其能夠生成多樣樣本並具備較高的覆蓋率，顯示在面對學生學習行為的不確定性時，VAE 具有更好的表示能力。**

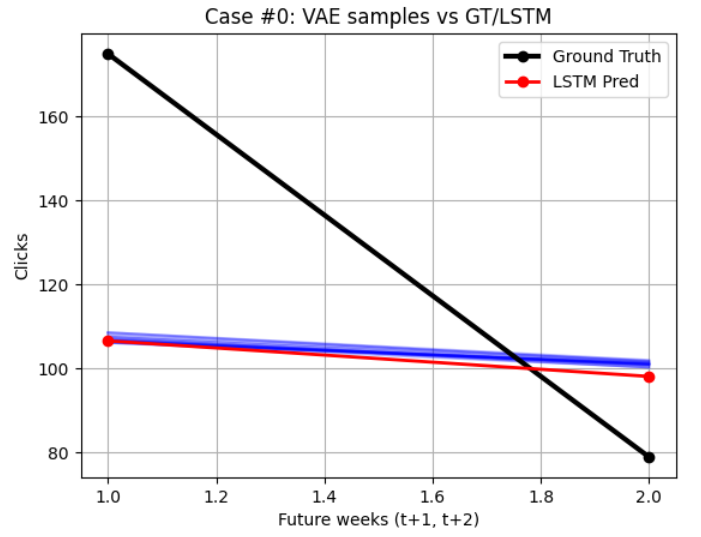
## 

## **視覺化結果**

* LSTM 模型：輸出單一路徑預測
  + 在 Sample #0、#1、#2 中可觀察到，LSTM 預測曲線（橘線）普遍較為平滑，無法反映真實行為的劇烈變化。例如 Sample #0 的真實點擊數急劇下降，但模型僅給出緩慢下滑的趨勢。



* VAE 模型：生成 N 條序列樣本（例如 20 條）
  + 在 Case #0 圖中，藍色半透明曲線為 20 條生成樣本，紅線為 LSTM 預測，黑線為真實值。可以看出，VAE 的生成範圍包覆真實值範圍，能有效呈現潛在行為的不確定性，提升覆蓋率與多樣性。



## 

## **優缺點討論**

| 模型 | 優點 | 缺點 |
| --- | --- | --- |
| Seq2Seq LSTM | * 結構簡單 * 訓練穩定 * 推論速度快 * 適合單一趨勢序列預測 | * 無法捕捉學習行為的突變或不確定性 * 預測結果過於平滑 |
| Seq2Seq VAE | * 能透過潛在變數生成多樣結果 * Best-of-N 表現更佳 * 覆蓋率高，能模擬多種未來情境 | * 訓練時間較長 * 需調整 β（KL loss 權重） * 可能出現生成樣本過於分散的問題 |

整體而言，LSTM 適合用於穩定、趨勢明確的預測任務，而 VAE 更適合應用於具有高度不確定性與多樣性的行為建模任務。在教育資料中，學生行為受到多重因素影響，VAE 提供了一種更具彈性的生成式建模途徑。