

以 LSTM 預測隨機亂數

B093040007 張碩文

摘要

在本專案中，會使用兩種生成隨機數的方式，分別為 LCG、linux /dev/random，生成好幾筆長度為 10000 的 0-9 之整數，讓 LSTM 以前面 5000 個 0-9 之亂數作為輸入，預測後 5000 個的 0-9 的整數，最後計算預測後的準確度，預期 LSTM 模型能學習到 LCG 生成隨機亂數的規則，而 linux /dev/random 的則無法學習其特徵，因為此隨機亂數本身是 truly random number，毫無規則性可言。

1 研究動機

老師上課時提到 LCG 的演算法可用類似差分方程的方式表達，且上課時有 demo 過，在大部分的角度時，雖然 LCG 生成的亂數看起來雜亂無章，但是從側邊的某個角度看，卻發現有很明顯的三個平面，因此好奇擅長處理時序型資料的 LSTM 是否能透過 LCG 所產生的隨機亂數作為訓練資料，訓練後學習到 LCG 生成亂數的方式，若能準確預測，意味著亂數生成的方式過於簡單，可讓模型輕易的學習到其特徵，同時也用 linux 提供的 /dev/random，其本身為 truly random number 作為對照組，因為本身的生成方式是從環境中的物理上的雜訊蒐集而成，無任何規則性，預期模型無法從 /dev/random 學習到任何規則。

2 方法

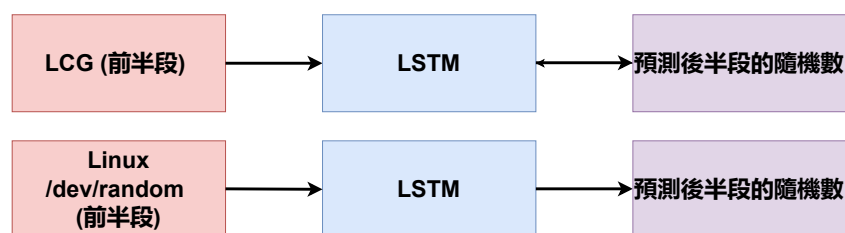


圖 1. 架構圖

如圖 1 所示，LSTM 模型的輸入為某隨機亂數序列的前半段，為 5000 個 0-9 之整數，模型預測為後半段 5000 個 0-9 之整數，且訓練 LCG 和 /dev/random 的模型架構完

全相同，僅差在模型內部訓練後的權重會不同，且架構會使用 pytorch 本身提供的一些 pre-trained 模型，再使用 tranfer learning 使得模型的權重訓練成更適合此資料集。爲了能清楚判斷是模型訓練結果不佳還是隨機亂數本身亂度夠高，會使用額外的亂數檢定法來分析亂數品質，像是 Monte Carlo value、Serial correlation coefficient、Entropy 等來衡量。

3 預期結果

預期模型能精準的預測 LCG 的隨機數，而 /dev/random 的則相反。

分類	項目	金額 (NTD)
戶外實境尋寶活動	尋寶遊戲設計與執行（含印刷費用）	\$120,000 - \$160,000
	工作人員薪資（時薪）	\$50,000
	影片宣傳和行銷	\$100,000
改建&活化建築	義村改造與設計	\$500,000
	藝術家設計&裝置藝術	\$250,000
	岡山老街建築整修	\$2,000,000 (optional)
螺絲博物館費用	道覽服務	\$20,000
	DIY體驗材料費	\$40,000
有獎徵答	宣傳和行銷	\$25,000
	購買獎項禮品	\$50,000
其他	保險費用	\$30,000
	雜費	\$30,000
總共費用		\$3,000,000