以 CNN‑LSTM 深度學習的奈及利亞、南非、波扎那 COVID‑19

感染病例預測模型

姓名：李采儒

指導老師：陳巧旻

摘要

新冠病毒疫情對全球社會，尤其是非洲國家造成了嚴峻的挑戰。這場疫情引發了經濟衰退、政治不穩定和大量失業等災難，對人們的生活和社會運作造成了嚴重的影響。疫情以驚人的速度蔓延，造成了許多生命的喪失。特別是在變異株的爆發之下，COVID-19感染病例的預測和確定變得至關重要，不僅對政府而言，也對整個社會而言。了解病毒傳播的情況可以幫助政府及時採取行動，制定新的政策和方法，以最大限度地減少COVID-19疫情對公共衛生和經濟領域的衝擊。因此此研究結合了卷積神經網絡（CNN）和長短期記憶（LSTM）的深度學習演算法，分別開發了奈及利亞、南非和波扎那的COVID-19感染病例預測模型。這些模型能夠根據過去的數據對未來10天的感染病例進行預測。

關鍵字：CNN、LSTM、CNN-LSTM、非洲、covid-19、模型預測

1. 研究動機與目的

冠狀病毒是一類感染和威脅哺乳動物和鳥類的病毒，包括MERS-CoV、SARS-CoV和2019-nCoV。隨著病毒不斷變異和演變，出現了多個新冠病毒變異株，其中一個是Omicron。Omicron變異株具有高度的傳染性和傳播風險，引起全球關注。全球多個國家報告了Omicron變異株的病例，其中非洲國家最為嚴重。非洲大陸已進行大量的疫苗接種，但仍面臨COVID-19疫情的挑戰。而各國也開發了多種預測和預測COVID-19疫情的方法，包括基於深度學習、統計和數學的方法。這些方法可以幫助衛生保健管理者和政策制定者制定適當的措施和干預措施，以應對疫情並保護公共健康。

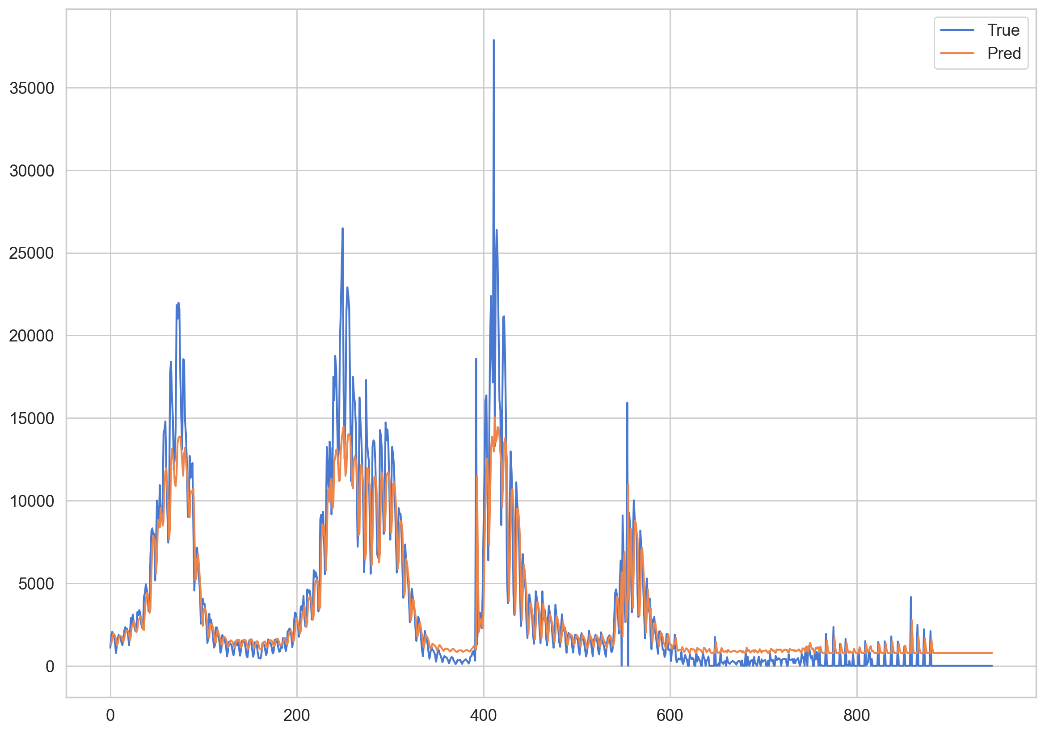
1. 論文研究內容

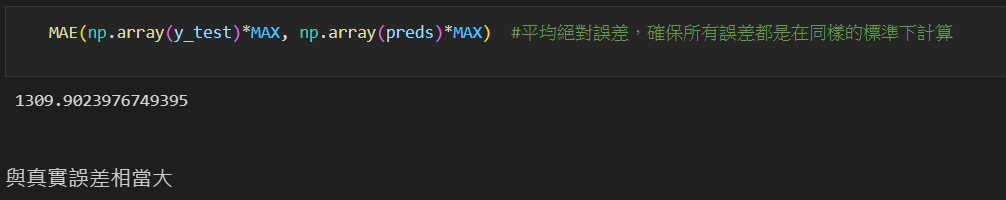
結合CNN與LSTM的模型進行預測。此研究使用一維CNN，由卷積層和池化層組成，對輸入數據進行數學處理，生成特徵。接著利用生成的特徵，使用LSTM和全連接層進行預測。此CNN-LSTM演算法將CNN作為編碼器，而LSTM作為解碼器。編碼器從輸入數據中學習特徵，並輸入到解碼器（LSTM）中。解碼器識別並建模數據集中內在的短期和長期時間關係。研究步驟簡要說明如下。

* + 1. 輸入層：接收輸入數據。
    2. 第一個卷積層：掃描上述i中的輸入信息，然後將結果應用於特徵圖。
    3. 第二個卷積層：再次找到特徵圖，以增強任何顯著特徵，每個卷積層有32個特徵圖，內核大小為3倍的步驟以讀取輸入序列。
    4. 最大池化層：從iii中刪除特定特徵，簡化特徵圖並生成具有較小維度的矩陣。
    5. Dropout層：增強學習網絡，防止過度擬合模型。
    6. 平坦層：將提煉的特徵圖展平為一個長向量，可作為解碼的輸入。
    7. 重複向量層：將輸入序列的內部表示重複，每個時間步長對應一次重複。
    8. LSTM解碼器：具有100個單元的隱藏層，能夠輸出完整的排列，每個單元每天提供一個值，作為預測下一天輸出序列的基礎。
    9. 全連接層：理解輸出序列中的每個步驟，以得到相似的層，為預測單個序列輸出提供支持，即LSTM解碼器能夠在任何時間進行操作，就像輸出層和全連接層一樣。
    10. 輸出層：預測新感染和死亡病例的新COVID-19病例數量，預測未來10天的情況。

1. 系統練習

使用同樣用來預測及分析資料的CNN-LSTM演算法，修改成適合用來進行分析及訓練，但效果有限，且尚未能進行未來十天的預測。

1. 結果

 圖一為從2020年1月至2023年5月進行訓練並預測的結果圖。

圖二為預測結果的平均絕對誤差。

1. 使用、參考資料

論文： <https://link.springer.com/article/10.1007/s12553-022-00711-5#article-info>

程式碼參考：<https://pseudo-lab.github.io/Tutorial-Book-en/chapters/en/time-series/intro.html>

數據來源：<https://github.com/owid/covid-19-data>