機器學習及應用期末報告

|  |  |
| --- | --- |
| 學號:112033640 | 姓名:謝佳峻 |

1. **介紹**

空氣汙染隨著工業發展越來越嚴重，近年來已成為一個嚴重的問題，空汙中的懸浮粒子(PM ,Particular Matter)被證明對人體有很大的影響，研究顯示暴露於充滿懸浮粒子的環境中與死亡率有關，近期已在全球引起了廣泛的關注。空氣汙染的預測對於空氣汙染防治與環境管理有重要的作用（Maleki et al., 2019），因此空氣汙染的預測已在研究界成為一項重要的議題。

空氣汙染的預測上會使用氣象因素(例如風向與濕度)和其他汙染因素(如PM)在先前相關歷史數據進行時間序列問題建模。根據相關研究的表示，空汙程度在時間上是一個動態且連續的過程(Maleki et al., 2019; Li et al., 2017; Feng et al., 2020) 。因此每一時刻空氣汙染的狀況都與前一刻相關，並會影響下一刻。因此在預測的方法上選擇能有效結合每個時刻關聯性尤其重要。

神經網路隨著硬體設備進步導致計算能力上升，進而推動了人工智慧與深度學習技術的迅速發展。在時間序列的預測上，長短期記憶網路LSTM及閘門循環單元(GRU)單元之間的連接沿著序列在時間步驟上傳遞訊息，使其能夠對時間序列展現動態時間行為。近年來兩者皆在時間序列的資料上展示出巨大的潛力，因此我希望利用LSTM及GRU預測空汙情形。

在本文中將使用雙向LSTM(Bi-LSTM)、雙向GRU(Bi-GRU)及殘差LSTM(Residual LSTM,KIM et al.,2017)進行空汙數值的預測，這三種模型的選擇基於它們在時間序列數據處理中的優越性能，尤其是在捕捉數據中的時間依賴性和非線性關係方面。

本文中其餘章節如下，第二節將介紹數據處理的流程，第三節將介紹上述三種神經網路的架構，第四節將介紹及分析實驗結果。

1. **數據處理**
   1. **數據分析**

|  |  |
| --- | --- |
| 一張含有 圖表, 螢幕擷取畫面, 方案, 設計 的圖片  自動產生的描述 | **一張含有 文字, 螢幕擷取畫面, 圖表, 數字 的圖片  自動產生的描述** |
| (a)氣象因素變數分佈 | (b)天氣分佈 |
| 圖(一) 訓練集數據分佈 | |

數據處理中首先進行數據分析，如圖(一)所示，在訓練及數據分佈中，訓練集變數種類共有8種，其中風向為文字描述，將其轉換成數值方法將在2.2章節介紹。天氣分佈在有無下雪的長條圖中，無下雪的數據佔絕大部分，為了避免輸入參數過多導致神經網路過擬合與計算負擔過重，訓練過程中將移除下雪天氣的數據。

* 1. **獨熱編碼(One-Hot encoding)**

獨熱編碼是一種將類別變數轉換為二進制向量的技術，每個類別的數值接用一個獨立的二進制變量表示，其中避免將變數直接轉換成數值所造成虛假的順序關係。訓練集中的風向為文字描述，為了能在訓練神經網路時作為數據使用，將透過獨熱編碼轉換成多個獨立的二進制變數，過程如圖(二)所示。

|  |
| --- |
|  |
| 圖(二) 獨熱編碼過程 |

* 1. **數據窗口化(Data Windowing)**

在處理時間序列的數據，會利用數據窗口化將長時間序列分割成較小片段(窗口)，每個窗口包含固定數量的連續數據，此項技術能夠捕捉局部數據的趨勢及特徵，以提升模型的預測能力。過程中首先會確認窗口大小，如圖(三)所示，選擇每個窗口包含多少數據，接續會滑動圖(四)，每次移動一個或多個數據點，產生重疊或不重疊的窗口。

|  |  |
| --- | --- |
|  |  |
| 圖(三)窗口大小設置 |  |
|  |  |
| 圖(四)滑動窗口 |  |

1. **方法**
   1. 雙向長短期記憶網路(Bi-LSTM)

雙向長短期記憶網路是一種改良的LSTM圖(五)，從圖中可以看到架構為前向及後向LSTM組成，這種架構能同時考慮雙向的時間序列特徵，從而提高模型對時間序列的理解能力。在本研究中所使用的架構如圖(六)所示，由數個雙向LSTM組成，每個雙向LSTM間接有使用全連接層使全部架構一致。

|  |
| --- |
| **Bidirectional LSTM (BiLSTM) Training System - GM-RKB** |
| 圖(五)雙向長短期記憶網路 |
| **一張含有 文字, 螢幕擷取畫面, 圖表, 字型 的圖片  自動產生的描述** |
| 圖(六)多層雙向長短期記憶網路架構 |

* 1. 雙向閘門循環單元(Bi-GRU)

閘門循環單元是一種RNN的變體，主要是由能控制先前的信息應該傳遞到當前時間的更新閘門及控制應忽視信息的重置閘門組成，其架構簡單使訓練成本低。雙向的優勢與LSTM一致，主要目的皆為可接受雙向的信息，就的架構如圖(八)所示，與雙向LSTM相似，差異在於將LSTM換成GRU。

|  |
| --- |
| **Structure of Bidirectional GRU (BiGRU): (a) GRU cell and (b) unroll BiGRU.  | Download Scientific Diagram** |
| 圖(七)雙向閘門循環單元 |
|  |
| 圖(八)多層雙向閘門循環單元架構 |

* 1. 殘差長短期記憶網路(Residual LSTM)

殘差長短期記憶網路結合LSTM與Resnet中殘差連接的設計，能有效提升網路的訓練效果和穩定性。殘差連接會在訓練時直接將信息傳遞至後面的LSTM層，如圖(九)所示，使的再訓練深層網路時能更好的訓練。

|  |
| --- |
|  |
| 圖(九)殘差長短期記憶網路層 |
|  |
| 圖(十)多層殘差長短期記憶網路架構 |

1. **實驗結果**

訓練時為了加快訓練速度，batch size設為32，窗口大小為11，迭代次數為150次，在訓練時由於資料皆在不同的尺度上，因此在訓練集皆有進行正規化，在表(一)中的RMSE及MAE的大小因此看似很小。

從訓練與驗證及的損失大小可以發現，Bi-GRU在高達150迭代大小下過擬合，但在程式中皆有設定自動儲存最低驗證誤差模型的演算法，因此在結果上可以避免訓練參數的一致，導致模型間的不公平比較。

|  |  |
| --- | --- |
|  |  |
| 圖(十一)Bi-LSTM 損失大小 | |
|  |  |
| 圖(十一)Bi-GRU 損失大小 | |

接續看到Residual LSTM(圖十二)，在驗證誤差上較前兩者穩定，與訓練集的趨勢幾乎一致，但看到表(一)，在相同層數與節點數量時，精準度皆輸給另外兩者，其中Bi-GRU有著最高的精準度。原因可能是由於Residual LSTM為單向結構，在獲取時間相關的信息下比另外兩者差，但在穩定性上最好，幾乎無過擬合情況的發生。

本次報告皆取出訓練時獲得的前三結果進行比較，訓練時有嘗試將窗口大小、增加層數及改變架構等嘗試，但結果皆不符合期待。

|  |  |
| --- | --- |
|  |  |
| 圖(十二)Residual LSTM 損失大小 | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| model | Layers × nodes | RMSE | MAE | R |
| Bi-LSTM | 16×3 | 0.064 | 0.038 | 0.591 |
| 32×3 | 0.061 | 0.037 | 0.627 |
| 16×5 | 0.062 | 0.038 | 0.610 |
| Bi-GRU | 16×3 | 0.058 | 0.036 | 0.664 |
| 32×3 | 0.061 | 0.040 | 0.616 |
| 16×5 | 0.065 | 0.040 | 0.576 |
| Residual  LSTM | 16×3 | 0.071 | 0.045 | 0.487 |
| 32×3 | 0.064 | 0.039 | 0.587 |
| 16×5 | 0.079 | 0.050 | 0.366 |
| 表(一) | | | | |

Reference:

Maleki, Heidar, et al. "Air pollution prediction by using an artificial neural network model." *Clean technologies and environmental policy* 21 (2019): 1341-1352.

Li, Xiang, et al. "Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation." *Environmental pollution* 231 (2017): 997-1004.

Feng, Rui, et al. "Analysis and accurate prediction of ambient PM2. 5 in China using Multi-layer Perceptron." *Atmospheric environment* 232 (2020): 117534.

https://www.linkedin.com/pulse/data-windowing-technique-used-time-series-forecasting-alejandro

https://www.gabormelli.com/RKB/Bidirectional\_LSTM\_%28BiLSTM%29\_Training\_System

https://www.researchgate.net/figure/Structure-of-Bidirectional-GRU-BiGRU-a-GRU-cell-and-b-unroll-BiGRU\_fig5\_360614707

https://www.kaggle.com/competitions/tlvmc-parkinsons-freezing-gait-prediction/discussion/416410

Kim, Jaeyoung, Mostafa El-Khamy, and Jungwon Lee. "Residual LSTM: Design of a deep recurrent architecture for distant speech recognition." arXiv preprint arXiv:1701.03360 (2017)