

天氣預報聊天機器人

專題編號：108-CSIE-S012

執行期限：107 年第 1 學期至 108 年第 1 學期

指導教授：王正豪

專題參與人員：105820006 江俊廷

105820020 梁丞鑫

105820024 廖勗宏

一、摘要

本系統使用中央氣象局長年來累積的龐大歷史數據為基礎，拋棄傳統上使用物理模型的天氣預報系統，以深度學習的方式進行數據分析及預測，並使用聊天機器人配合 LINE，給予使用者最方便的介面、最即時的反饋。

關鍵詞：weather forecast, data mining, deep learning, time series, chatbot

二、緣由與目的

傳統上，天氣預報是通過大氣的物理模型來完成的，但些微的擾動就可能使系統不穩定，因此只要欲預測的時間較遠，其結果就會變得非常不準確。

機器學習技術在擾動的影響方面相對減少許多，因此探討其在水天氣預報中的潛能及應用，以生成更準確且更長時間的天氣預報。此外，深度學習相較於傳統的統計模型，不需要頻繁地更新模型。

目前在市面上並沒有互動式的氣象界面，因此我們將氣象預報結合 LINE 聊天機器人，增加互動、貼近社群、並提升多平台的通用性。

三、研發工具

資料庫：MongoDB

後端：Flask

機器學習：Scikit-Learn, Keras, Tensorflow

UI 介面：Line Bot Designer, Dialogflow

四、系統架構與方法

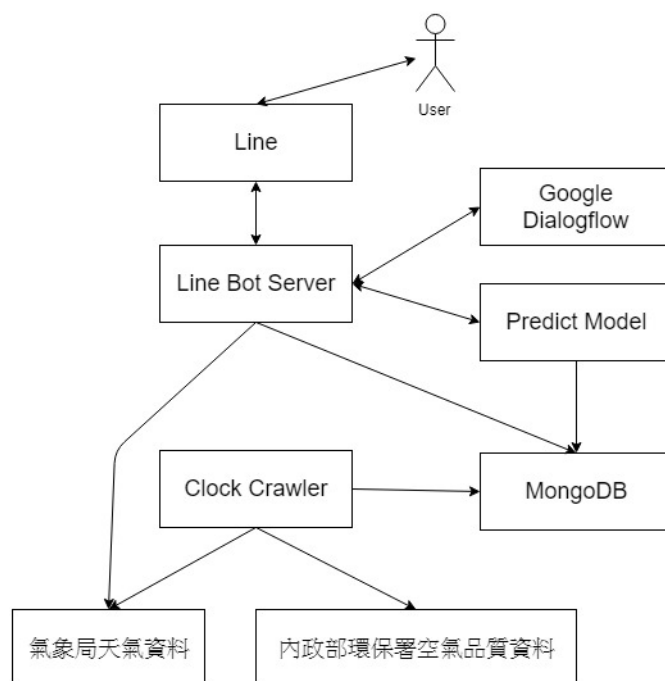


圖 1. 系統架構圖

方法：

1. 預測模型：

在資料前處理上，將缺失資料填入近日平均值後作正規化。

溫度預測：將可預測的時間範圍加長至一週，建立所有測站皆可使用的整合模型。

雨量預測：利用短期的雨量資料加上基礎的神經網絡，預測當日內的降雨氣象變化。

PM2.5 空氣品質監測：依據半日內各項空氣品質數據，預測 4 小時後的空氣品質狀況，並利用 LSTM 提升預測極值以及時間序列的準確性。

最後，將訓練好的模型使用交叉驗證

以及 Loss function (RMSE)進行評估

2. 資料庫：

儲存中央氣象局以及行政院環境保護署的開放數據和預測模型的預測結果，並定期更新。

另外，當新的觀測資料寫入資料庫時，資料庫會比對預測資料與觀測資料，評估誤差大小是否需重新訓練預測模型。

3. 聊天機器人：

使用 Dialogflow 為聊天機器人進行判讀關鍵字詞及意圖的訓練，完成後使其與 Python Flask 串接。

在問句設計上採用關聯式問句以及 LINE 的回答選項，當使用者與聊天機器人互動，經由 dialogflow 判斷後再連接後端及資料庫，根據問題做不同回應。

五、成果

使用者透過 LINE 接收豪大雨以及空氣品質下降等較須注意的提醒，且可向聊天機器人詢問降雨機率、溫度、空氣品質等天氣狀況。

準確率：

	溫度	雨量	PM2.5
RMSE	2.203	43.272	9.054
Variance	0.88	0.78	0.73

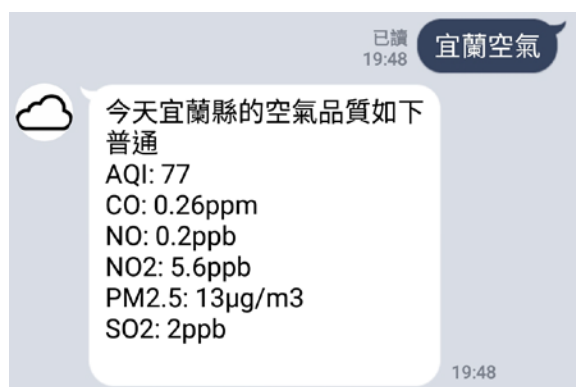


圖 2. 天氣功能描述

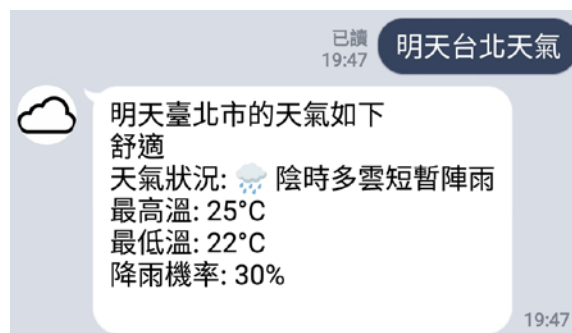


圖 3. 空氣品質功能描述

參考文獻

- [1] Nikolay Laptev, Slawek Smyl, and Santhosh Shanmugam, "Engineering Extreme Event Forecasting at Uber with Recurrent Neural Networks" June 9, 2017
- [2] Chiou-Jye Huang, "A Deep CNN-LSTM Model for Particulate Matter (PM2.5) Forecasting in Smart Cities", *Jiangxi University of Science and Technology*, July, 2018
- [3] Tomoaki Kashiwaoa, Koichi Nakayamaa, Shin Andoc, Kenji Ikeda d, Moonyong Lee, Alireza Bahadori, "A neural network-based local rainfall prediction system using meteorological data on the Internet: A case study using data from the Japan Meteorological Agency", March 18, 2017