

機器學習與天氣預報系統

專題編號：108-CSIE-S008-MID

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

本系統以機器學習技術，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，以人工智慧的方式進行數據分析及預測。

關鍵詞：weather forecast, data mining, regression, decision tree, machine learning

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。

三、研發工具

網頁系統：Flask, Vue.js, MongoDB

爬取資料：Python

機器學習：Scikit-Learn

(一)MongoDB

因為後端使用 Python，而 MongoDB 內的資料是使用 JSON 格式儲存，可以直接轉成 Python 的字典，不需要另外處理。

另外，氣象和雨量資料寫入資料量會很

大，且對於簡單的查詢需要快速的回應，因此採用非關聯式資料庫。

(二) Flask

Flask 比較輕量化，我們的系統在後端不會有太多的功能需求，因此用輕量化的 Flask 再擴充即可。

(三)Task scheduler

使用 python 所寫，利用作業系統課程中所學的 thread，可定期抓取氣象局資料。

四、系統架構與方法

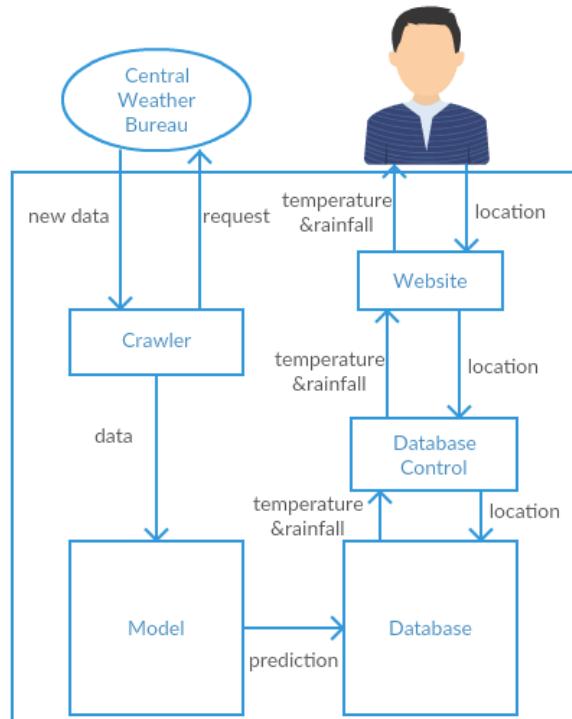


圖1. 系統架構圖

(一)溫度預測：regression

為找出溫度與各天氣屬性的關係進而預測，因此使用迴歸進行分析，找出能夠代表觀測資料的函數，分析資料如下：

1. data cleaning:

利用 python pandas 選取資料完整之屬性，文字資料轉換為數字，遺漏值填入數值。

2. Feature select:

先由 correlation 判斷(0.6以上)，使用 backward elimination 逐步刪除。

利用 python statsmodels 建立 OLS 模型，利用 t-test($p < 0.05$)找出統計顯著之值。

3. 訓練模型：

使用 python scikit-learn 套件正規化後，分為訓練與測試集，建構迴歸模型，得出準確率。

4. 模型評估：

使用 cross validation 得出準度均值及標準差判斷模型是否過擬合。

(二)雨量預測：decision tree

1. Data cleaning:

優先刪掉大量空白的欄位後，將能夠從原資料獲取的欄位刪除，對於資料缺失的部分欄位，用兩種方法測試：填0和填入平均值。

2. Feature choose:

在選擇的 Feature 中，有額外增加兩個欄位：一天後的雨量資料以及兩天後的雨量資料。使用原生的資料當作輸入欄位，額外增加的雨量資料當作預測欄位下去做模型建立。

3. 訓練模型：

使用 python scikit-learn 套件，將資料訓練和測試集，建構決策樹。

五、進度與未來規劃

(一)溫度預測

實驗數據使用中央氣象局開放資料(台北)，時間為2017年每小時一筆，由前3小時的資料去預測結果。

最後選取溫度與氣壓幾項屬性，預測準度為0.94左右，經由 cross validation 測試有過擬合傾向，所以使用 data shuffle 避免，效果有所提升。

(二)雨量預測

預測的資料是以一天的雨量當作一筆資料，使用約10年的資料當作資料庫，但由於雨量是一個極度不穩定的資料，極端值很大且大多數時間數據皆是0，所以預測出來的模型目前為止並不優秀，有待修改。

(三)未來規劃

1. 雨量預測方法的調整：

預計會使用深度學習技術。

2. 增加訓練模型所使用的資料量：

輸入更多年份和屬性的資料進行測試。

3. 網頁畫面視覺化以及操作直覺化：

以台灣行政區圖供使用者點選。

參考文獻

- [1] Mark Holmstrom, Dylan Liu, Christopher Vo, “Machine Learning Applied to Weather Forecasting” *Stanford University*, December 15, 2016
- [2] Adam McQuistan, “Using Machine Learning to Predict the Weather: Part 1”, November 14, 2017
- [3] Siddharth S. Bhatkande, Roopa G. Hubballi, “Weather Prediction Based on Decision Tree Algorithm Using Data Mining Techniques”, *Department of CSE, KLE Dr. MSS College of Engg & Tech. Belgaum India*, May, 2016