

1. Preprocessing/Feature Engineering

我處理 feature 的方式主要核心是利用前 10 周的 feature 和 label

來預測第 11 周的結果，而我抽取的 feature 分別是：

「reanalysis_dew_point_temp_k」，

「reanalysis_specific_humidity_g_per_kg」，

「station_min_temp_c」，「station_avg_temp_c」以及上面列舉這

4 項各別的三次方，所以對於每周來說，我取的 feature 共有上面

8 個以及每周的登革熱案例數目，總共有 9 筆 feature，因此，每

10 周為單位來取的話就相當於一筆完整預測的 feature 共有

$9 \times 10 = 90$ 筆資料，而在從 training set 取得很多上述型態的 90 筆

feature set 後，會將這些 data 都標準化。

總結來說，我對 feature 的抽取就是利用上述方式獲得前 10 周的

90 筆資料後，利用這些 feature 來預測第 11 周的登革熱案例數。

然而，針對 testing set 的部分，因為它的第一周剛好對應到

training set 結束後的下一周，因此我的做法就是將 training set 的

最後 10 周抽出來的 feature 用來預測 testing set 第一周的案例數，

接著就將 training set 的最後九周加上 testing set 第一周的 feature

來預測 testing set 第二周的案例數，以此類推。

2. Model Description

我的 model 是利用 RNN 和 Random forest ensemble 預測而成的，

RNN 的架構請見另外一則介紹，而 Random forest 的部分就是利用每筆 feature set 共 90 個 feature 當作 training input，分成預測 sj city 和 iq city 的 model，sj city 的 random forest model 是利用 250 個 tree 來合併預測，iq city 的部份是利用 200 個 tree 合併預測。在獲得 validation accuracy 不錯的 model 後，我們將 RNN 的 model 預測出來的案例數和 Random forest 預測出來的案例數以 8:2 的比例作 ensemble 來得到 testing set 的預測值。