

如何奪得先機搶發PTT地震爆文

Motivation

5/3 11:29 我在臺北的社科院感受到地震，幾秒後我突然好奇，若在這時打開PTT的八卦版，會看到幾篇地震文。結果已然有十數篇，更發現第一篇早已推爆。且發文時間在11:28:18。這意味著早在我感受到地震以前，在別處已有人感受到地震而發文。這不但受地震的傳播速度影響，更讓我發想了以下的問題：究竟要怎樣才能奪得先機搶發PTT地震爆文呢？要怎樣才能如推文所說的，真正地在發地震爆文上「贏了」呢？

問題發想

根據科學月刊的文章，P波每秒走5-7 km，S波每秒走3-4 km。而5/3上午的地震震央在台東臺北方海域，爆文則由來自台中的網友奪得。上一次地震則是4/12，震央及爆文發文者均在宜蘭。因此可以提出以下的問題：

- PTT使用者與震央的遠近是否會顯著地影響奪得爆文的機率

與此同時，我們必須進一步問的可能就是，如果會，那麼：

- 如果我不在震央所在的縣市，我應該要快幾秒才可以搶到地震文？

為了要回答這個問題，我們顯然需要發文者的位置，幸好PTT已經紀錄了IP位址，我們只要反查即可。如：<https://www.ez2o.com/App/Net/IP>

以及利用中央氣象局發佈的地震報告當中的震央資訊，即可比對發文者位置與震央位址的遠近。

其他控制變數

其他可能影響使用者搶發地震文的因素有：

1. 地震發生時間 在深夜或者凌晨的地震很可能沒有多少使用者在線上，所以會影響搶發地震爆文的機率
2. 地震震央是否靠近大都市 猜想PTT使用者多數聚集在人口較多的城市地區，因此當地震發生位置在人口密度較低的縣市附近時，位在該縣市的使用者搶得爆文的機率可能沒有比「當地震發生在人口密度高的縣市附近時」來得高。意思就是：住在人口密度較低的縣市的PTT使用者比較沒機會搶到爆文
3. Selection Bias 沒搶到爆文的其他地震文會被刪掉，如果可以找回這些文章，也看看他們的IP位置，就可以還原回各地區搶發地震文的時間落差。

流程

1. 爬PTT Gossip版上標題為「地震」的文章，同時紀錄IP位址
2. 爬中央氣象局的地震報告，將震度超過一定程度的篩選出來。同時與PTT文章的日期對應，有日期一致的則分類至同一起地震。
3. 將PTT文章的IP餵給查詢IP的網站，得到位置訊息後紀錄下來。將1~3的資料整理成Dataframe
4. 將其他控制變數也整合進Dataframe裡，如發文時間

Data

每週計畫進度

- 5/26：爬蟲 - PTT地震文、氣象局地震報告
- 6/2：清理資料、跑迴歸
- 6/9：地理資訊視覺化、撰寫報告