**北大 x 好域**

時間：2019/12/19（四） 16:00

參與人：陳宗天教授、林奕銜、黃俊友、蕭琮峻、許程閔、溫智皓

上次[會議紀錄](https://docs.google.com/document/d/1QGalFfvxlEHqv2hD0rlRx9BTXH3MHFTRA_Szj-7jNig/edit)，[Working Process](https://docs.google.com/document/d/1AUf4gI5kxPYSEiniz-MhMeHtMWicBmLqcMjot6EHihM/edit?usp=sharing)

**問題討論：**

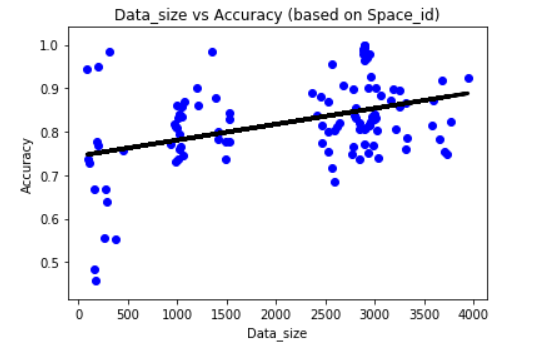
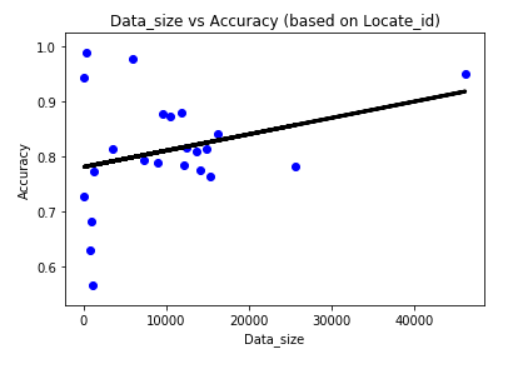
1. 因使用率（一般及月票數量比例約 1：1）關係，是否要將月票訂單加入模型考量在內？
   1. 由於月票性質和預測模型的動態定價目標性質不太相同，於預測模型上，我們認為只使用一般訂單才能較為準確達成我們的目標。
   2. 而月票訂單部分，目前則是傾向於做相關的統計分析，作為未來決策的依據。例如：同一空間的一般訂單較少，而幾乎都是月票用戶預定，代表該空間可能訂價太高。
2. 是否需縮小資料比重差異（真假訂單目前比率約為 3：7），讓假訂單時間在 1 ~ 4時段隨機產生？
   1. 除了真實訂單使用到的時間，為了反映實際空間使用情形，應讓假訂單填充空閒時段，才能準確預測真實情形時，該時段被出租的機率。
3. 如果相關性高可以去掉其中一個沒問題，如果Price、Size都去掉，是否可能讓相同locate的space彼此之間沒有什麼差異？
   1. 最後決定：根據模型評估績效，去掉price，留下price\_per\_size、size。
4. 如果利用space準確率也夠高，是否可以考慮使用space，因為space才可以準確反映每個空間的特色（大小、單價、擁有的設備...等），未來新空間的加入可能也比較好利用現有模型？單用locate很難有依據去解釋每個locate之間要怎麼做連結
   1. 雖房間一模一樣，實際在現場的情況會有所差異（如是否靠窗、靠廁所等），或許未來可再將這些資訊放進模型。
   2. 模型比較：總模型 vs locate vs space

|  |  |  |  |
| --- | --- | --- | --- |
|  | 總模型 | locate | space |
| 準確率 | 中等 | 變動中等 | 變動大 |
| 資料量 | 大 | 中 | 小 |
| 泛用性 | 廣泛 | 中 | 單一 |

1. 現在做的模型必須考慮之後新模型加入時要怎麼把新空間與現有空間做連結，畢竟舊的空間有租約到期問題，不能一直沿用，模型必須是能套用在所有locate上的，如果僅看獨立模型，可能較難以一直使用。
   1. 模型是根據locate內space的設備、價格來判斷，未來若有相類似的空間產生，就能以舊空間所訓練出來的模型來做預測。
   2. 未來可嘗試使用cluster或其他分析方式來比較不同locate的相似性，進而選擇最相似的來進行預測。

**目前進度：**

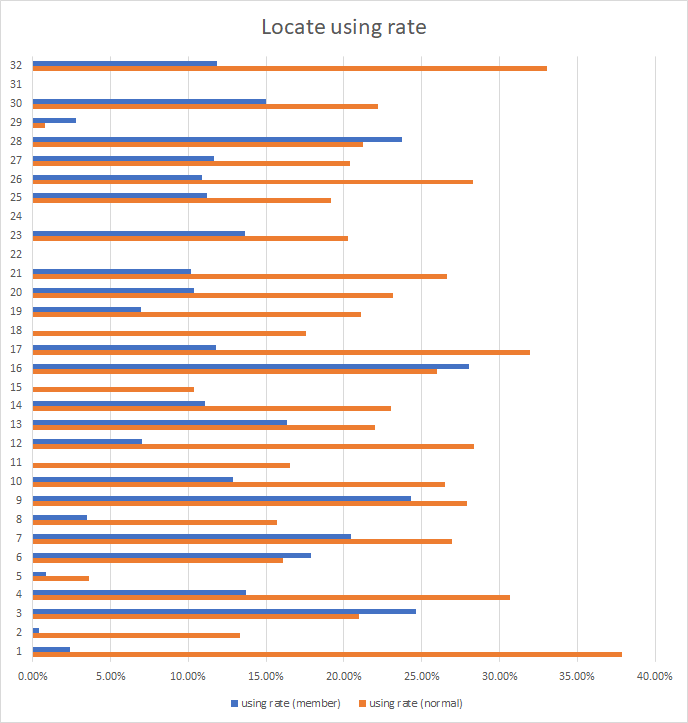
1. 模型準確率、data量關係：



（圖一） (圖二)

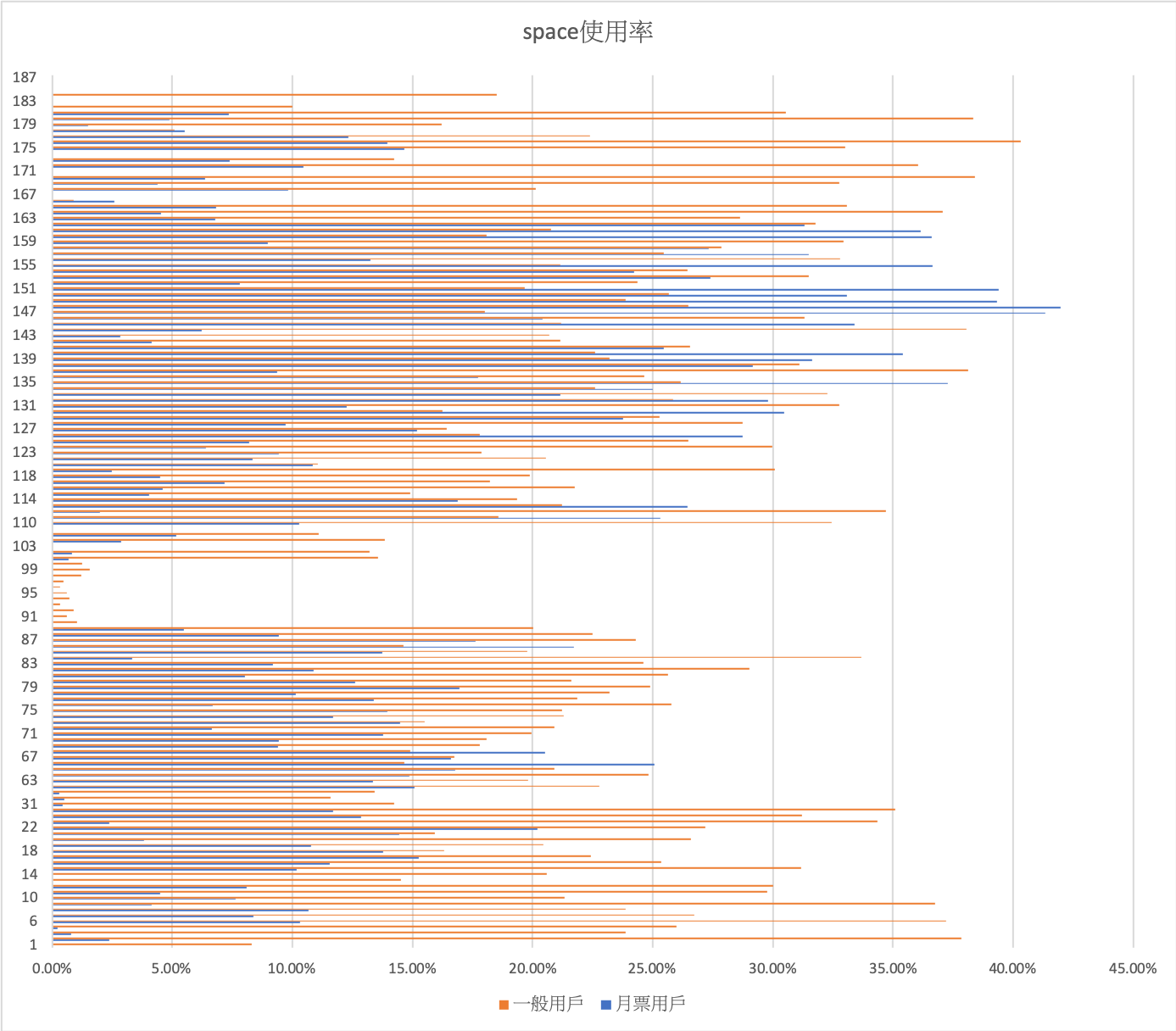
從上述（圖一）及（圖二）目前準確率和資料量有正相關，可以看出資料量多的預測結果較準確，也已提高分別模型的準確率。

1. 統計分析
   1. [locate](https://docs.google.com/spreadsheets/d/1lhefIJwPJr5iTsyb-v8OAQonxmsaOxl6ZkjzNChxH2A/edit?usp=sharing) (CSV檔案)
      1. 整體locate一般用戶平均使用率：19.74%
      2. 整體locate月票用戶平均使用率：10.11%



（圖三）

* 1. [space](https://docs.google.com/spreadsheets/d/12i0EbbKAznZa_uSM7KQVuqMIqj_hzofmFJTHxr9oj4c/edit?usp=sharing) (CSV檔案)
     1. 整體space一般用戶平均使用率: 20.58%
     2. 整體space月票用戶平均使用率:



（圖四 [細部圖](https://drive.google.com/file/d/16SnEzpgswnESogH9d_1Ic5mrcj8IbbOu/view)）

從上述（圖三）、（圖四）皆可看出，每個不同locate、space的月票用戶及一般用戶的使用比例上都有很大的差異，有些空間幾乎只有一般月戶，而有些空間則是月票用戶居多。未來這些分析數值可提供決策參考，另外可做為調整預測模型準確率的參考之一。

**下一個目標 & 未來可嘗試的事情：**

1. 資料處理
   1. 不同locate的獨立模型，準確率會有很大的差異，找出影響準確率的因素（真訂單佔比、不同時段佔比等等）
   2. 瞭解月票使用者（月票使用者id）對locate、space的影響與關係。
   3. 結構化資料（現有訂單資料加以統計分析）
   4. 非結構化資料（文字、圖像資料等）
   5. 希望能拿到2019.11.18至今天的最新資料。