

國立成功大學  
工業與資訊管理學系碩士班  
碩士論文

公共自行車共享系統之群眾運補策略數學模型與資料視  
覺化工具設計研究

The Design of Mathematical Models and Data Visualization  
Tools for Public Bike Sharing Systems with Crowdsourced  
Repositioning Strategy

研究生：侯貞泰

指導教授：王逸琳 教授

中華民國一百零五年八月

國立成功大學

碩士論文

公共自行車共享系統之群眾運補策略數學模型與資料  
視覺化工具設計研究

The Design of Mathematical Models and Data  
Visualization Tools for Public Bike Sharing  
Systems with Crowdsourced Repositioning  
Strategy

研究生：侯貞泰

本論文業經審查及口試合格特此證明

論文考試委員：

王逸琳 李雨青

李寧波 王中允

指導教授：王逸琳

系(所)主管：李洪毅

中 華 民 國 105 年 5 月 26 日

## 摘要

近年來，公共自行車共享系統已在多個國家的都會區營運，也帶動了新型態的物流研究議題。目前此類系統在營運期間最棘手的挑戰為「無車可借」或「無位可還」的供需失衡問題，現行作法大多雇用卡車，自較多車的租借站移車，再將這些車載至較少車的租借站補車，然而少量的卡車無法滿足大量租借站同時缺車或缺位的運補需求。因此，本研究提出「群眾運補」的概念，在給予適當獎勵的情況下，由系統招募志願接受運補任務的群眾，這些群眾將依系統指示於特定時刻至指定的租借站租車，再騎至指定的租借站還車，如此即可同時在多個租借站間靠群眾運補，不會受到少數卡車所在位置及容量的限制，理應比現行使用少數卡車的運補方式更為有效。

本研究分成三個階段：在第一階段的資料分析中，我們以台北市 YouBike 系統的公開資料為基礎，藉由分析該資料找出租還需求的變化趨勢（譬如各站每日租借量隨時間的波動方式，以及其可能還車的訖點分佈方式等），並實作多種資料視覺化工具以呈現分析結果；第二階段探討群眾運補的數學理論基礎，建構一個理想的最適車數整數規劃模式，將一天分成多期，以計算出各租借站於各期期末的最佳庫存之自行車數量，該數值可被視為該站該時刻的最適車數，再以即時租借資料與機器學習中的隨機森林預測模型來預測需求，建構一個簡化的群眾運補數學模式，以計算當期最佳的群眾號召方式；第三階段則實作一個模擬系統，模擬以群眾（或卡車）運補的情境下，如何配置這些騎乘者（或卡車）的起訖運補方式，以減少最多的預估缺車與缺位總量；經多次模擬測試，證實群眾運補的確比傳統卡車運補有更高的效益，尤其在尖峰時段更能大幅減少缺車與缺位發生的機會，可明顯地改善服務品質。

**關鍵字：**自行車共享、群眾運補、資料視覺化、整數規劃、隨機森林、模擬

# **The Design of Mathematical Models and Data Visualization Tools for Public Bike Sharing Systems with Crowdsourced Repositioning Strategy**

Chen-Tai Hou

I-Lin Wang

Department of Industrial and Information Management

## **SUMMARY**

The public bike sharing system is a popular shared economy application. Its success depends on the availability of bikes and empty racks in rental sites spread out a metropolitan area with sufficient density. Its major management challenge is to reposition bikes between rental sites to prevent full or empty rental sites. To this end, most systems hire trucks to move bikes between rental sites, yet this repositioning scheme is ineffective because the number of trucks are usually much fewer than rental sites. Here we proposed a crowdsourced repositioning scheme, where voluntary riders are hired to ride specific Origin-Destination trips. Receiving satisfactory bonus in return, the voluntary crowd can move bikes at all rental sites simultaneously with less costs.

We conduct in-depth analysis on historical rental data to identify important factors and features affecting demands. Some visualization tools are designed to provide better understanding to the demand dynamics. The random forest algorithm is employed to predict the expected incoming bike rentals and returns. Setting every 30 mins as a period, we propose an Ideal Inventory Model (IIM) using integer program, and calculate optimal bike inventories for all rental sites and periods. To deal with on-line repositioning, we propose the Voluntary Rider Flow Model for period  $t$  (VRFM( $t$ )) that calculates voluntary rider flows by a greedy algorithm. The results of our tests by simulation indicate our crowdsourced repositioning strategy does provide better service quality than the truck repositioning strategy.

**Key words:** Bike sharing, Crowdsourced repositioning, Data visualization, Integer program, Random forest, Simulation

## INTRODUCTION

The recent boom in the sharing economy has led to new business and changed how people live in several ways. For example, the public bike sharing systems (PBSS) in many metropolitan areas are perfect to serve for the first and last mile connection to public transportation systems such as MRTs, trains, or buses. The main challenge of bike sharing systems is to dynamically reposition bikes between rental sites to keep good balanced bike inventory for each rental site at any time. Conventional dynamic repositioning strategies usually employ trucks to move bikes between rental sites. Previous dynamic repositioning literatures focus on integer programming approaches or heuristics to calculate optimal truck routes and bike loading/unloading schedules. Yet these approaches are not effective due to the following reasons: (1) the input parameters that measure the demand characteristics are usually not accurate for constructing the integer programs on a time-space network; (2) the demands are assumed to be static for integer programs, yet they should vary dynamically to some extent; (3) these integer programs involve too many variables and constraints and usually could not deal with cases of more than 60 rental sites, due to the complexity issues; and (4) the number of trucks is much smaller than the number of rental sites, which means most rental sites cannot be served at the same time.

To overcome the above drawbacks, this thesis proposes a novel crowdsourced repositioning strategy, where we recruit voluntary riders instead of trucks to reposition bikes. Assuming we can always find voluntary riders anywhere anytime by offering satisfactory compensation such as free riding hours, bonus, or coupons, this voluntary crowd can simultaneous repositioning bikes along specific Origin-Destination trips by following specific instructions determined by our models. To make more accurate demand forecasts, we use the random forest algorithm in machine learning to predict incoming bike rentals and returns based on real-time statistics. Based on similar time-space network settings but assuming unlimited voluntary riders, we first propose an integer programming formulation called the Ideal Inventory Model (IIM) to calculate the ideal bike inventories for each period. In some sense, this ideal bike inventories represent the optimal bike deployments in all rental sites at each period, and serve as a target bike inventory to achieve for a reduced version of IIM called the Voluntary Rider Flow Model at period  $t$  ( $VRFM(t)$ ) that determines optimal voluntary rider flows for all rental sites at period  $t$ . Since our approach considers more accurate demand data and determines the on-line repositioning decisions at each period, it is more accurate and flexible than conventional truck repositioning strategy.

## MATERIALS AND METHODS

To simplify the problem, we first use 30 min as the length of a period (e.g., 06:00 - 06:30 - 01:00 - 01:30 - ... - 23:30 - 24:00) to construct a time-space network. We determine the ideal bike inventory for each rental site in each period by the Ideal Inventory Model (IIM), which is an integer program over this time-space network, using the average number of bike rentals and returns in each period as inputs and unlimited available voluntary riders to move bikes from each site to its neighbor sites within one period. The IIM is much easier to solve than conventional dynamic repositioning integer programming models since there is no truck routes involved. The bike inventory at each rental site and period in an optimal IIM solution can be used as a target optimal bike inventory to be achieved for each site and period in the on-line crowdsourced repositioning model VRFM(t).

The integer program VRFM(t) in fact corresponds to a minimum cost flow problem. Here we design a greedy algorithm for VRFM(t) as follows: In the beginning of period t, for each rental site we estimate the number of bikes to be repositioned by the difference of its current real-time inventory and target inventory calculated by IIM. Each site is surplus, balanced, or deficit, depending on the difference to be positive, zero, or negative. Then, we repeat assigning maximal voluntary riders to move bikes from a most surplus site to a most deficit site, until no more riders can be assigned.

To make our VRFM(t) more effective, more accurate demand forecasts are required. To this end, we conduct in-depth analysis on the historical rental data to identify important factors and features affecting bike rentals and returns. We use a 10-month open rental data of the YouBike, the PBSS at Taipei, Taiwan, for our analysis. We design several visualization tools to help understand the rental patterns for all sites within one day. The derived data is also used as input for the random forest algorithm to help predict expected number of bike rentals and returns in current period.

A simulation scheme is developed for evaluating the performance of different repositioning strategies. We generate daily rental demands from the real PBSS rental data, either by importing the data from a specific day, or by mixing and sampling the data of several random days. In the beginning of every period t, we decide how to reposition bikes either by the greedy algorithm for VRFM(t) using the crowdsourcing strategy, or by another greedy algorithm for assigning trucks using the truck strategy. For the truck strategy, we first conduct a K-means clustering algorithm to partition the rental sites into disjoint service zones, so that each service zone has similar amount of total net rental data (i.e. the difference

between the total checking outs and returns) and the rental sites in one service zone are in close vicinity to each other.

## RESULTS AND DISCUSSION

We use 10-month historical YouBike rental data in 2014 as the testing data. We use sunny weekdays to calculate the idea inventory. Then we use weather data and historical data to train the random forest model. Based on our testing, we have selected seven important features that affect the prediction: holiday or not, weekday or not, day (i.e., Monday, ..., Sunday), current hour, number of checking out and returns in previous period, temperature, and rainfall. For each site, we construct its own random forest, which can output a predicted value on the number of checking outs and returns with the given seven parameters.

Based on our simulation testing, the crowdsourced repositioning strategy has improved up to 1-3.5% of unmet demands than repositioning by 10 to 25 trucks for the 18 actual daily rental data. Moreover, if we compare the effectiveness in the rush hours (e.g., 6pm), the crowdsourced repositioning may reduce up to 32.5% unmet demands than the truck repositioning. These results indicate the crowdsourced repositioning provides much better service than trucks, especially in rush hours. We also learn that the effectiveness of prediction by the random forest algorithm is around 0.8%. These results indicate the more accurate prediction (e.g., by the random forest algorithm) does improve the service quality.

## CONCLUSION

In order to reduce the unmet rental demands in the bike sharing systems, we propose a novel crowdsourced repositioning scheme and show that it is more effective than current truck repositioning strategy by simulations. Although similar idea has been mentioned for years, our work is arguably the first one to give detailed mathematical models and numerical experiments on how to implement it, to the best of our knowledge. The results of our computational experiments show that crowdsourced repositioning is a feasible solution and better than traditional truck repositioning. This method can solve the rush hour needs, and might arguably be the only and best way to reposition shared vehicles of larger sizes (e.g., cars or motorcycles). For future research, we suggest to investigate more accurate prediction models on rental demands, as well as marketing strategies in encouraging the voluntary riders.

## 誌謝

加入 ilin lab 到現在兩年半的時間，在這間研究室擁有了許多難忘的回憶，隨著碩士論文的完成，也代表著學生生涯即將告一個段落，感謝家人的支持，讓我能順利完成我的碩士學位。

感謝指導教授王逸琳老師的指導，從大四下到現在，一路參與公共自行車共享系統的研究、火車鐵路比賽、救災比賽以及最後的碩士論文研究，很感謝老師願意花時間指導我不足的地方，在每次 meeting 的過程中讓我學習到老師在面對問題時的思考方式，也漸漸的影響我在遇到問題時該如何處理。也很感謝老師知道我所感興趣的議題，讓我嘗試用例如視覺化、網站呈現來處理問題。有這兩年半的訓練，我相信對自己未來有很大的幫助。

感謝實驗室的夥伴，剛進實驗室就受到琮閔、阿佰、偉德的照顧，讓我很快的融入這個實驗室。在剛開始寫論文時尤其感謝阿佰的協助，幫忙檢查論文的內容。同一屆的小秉跟采緹，我們一起經歷過課業上、比賽上的崩潰，一起打屁、取暖、互相學習，希望夜鷹部隊以後都能變成早鳥成員 XD 還有富元、筱昀、BK、思涵、冠緯，感謝你們在寫論文煩悶時，陪打桌遊以及各種哈拉，如果未來有需要我協助的地方可以找我討論哦。

最後謝謝自己這兩年半來的努力，這些日子雖然很辛苦，但回過頭來這一切辛苦都是值得，也算是有達到自己當初念研究所時的期許。在成大這六年，難以用這短短的幾句話表達，感謝在這一路上遇到的師長、朋友，我會永遠珍惜這些回憶。

# 目錄

摘要 .....	I
誌謝 .....	VI
目錄 .....	VII
表目錄 .....	X
圖目錄 .....	XI
第一章 緒論 .....	1
1.1 研究背景 .....	1
1.2 研究動機 .....	2
1.3 研究目的 .....	4
1.4 論文架構 .....	6
第二章 文獻回顧 .....	8
2.1 運補策略之相關文獻 .....	8
2.1.1 靜態運補 .....	8
2.1.2 動態運補 .....	9
2.2 資料分析之相關文獻 .....	11
2.3 營運策略相關文獻 .....	15
2.4 隨機森林演算法 .....	15
2.5 小結 .....	16
第三章 資料分析與視覺化工具 .....	17
3.1 資料簡介 .....	17
3.2 資料分析 .....	18

3.2.1	平日與假日交易趨勢 .....	18
3.2.2	晴天與雨天 .....	20
3.2.3	起訖點趨勢 .....	20
3.2.4	時間區間 .....	21
3.3	視覺化網頁工具 .....	22
3.3.1	使用技術 .....	22
3.3.2	工具操作說明 .....	23
3.3.3	資訊呈現 .....	24
3.3.4	視覺化網頁工具小結 .....	29
3.4	小結 .....	29
<b>第四章 群眾運補之研究 .....</b>	<b>31</b>	
4.1	最適車數數學模式 .....	31
4.2	預測未來需求 .....	36
4.3	當期群眾運補數學模式 .....	37
4.3.1	當期群眾運補策略 .....	37
4.3.2	整數規劃模式 .....	39
4.4	小結 .....	41
<b>第五章 實際範例與數值分析 .....</b>	<b>42</b>	
5.1	測試情境 .....	42
5.2	最適車數數學模式 .....	43
5.3	隨機森林預測模型 .....	44
5.4	模擬測試 .....	48
5.5	小結 .....	57
<b>第六章 結論與未來研究方向建議 .....</b>	<b>59</b>	

6.1	結論與貢獻.....	59
6.2	建議之未來研究方向.....	62
<b>參考文獻 .....</b>		<b>65</b>
<b>附錄 A、模擬數值結果.....</b>		<b>68</b>



## 表目錄

表 1.1 以卡車運補之粗估參數範例	3
表 3.1 租借資料範例	17
表 3.2 中央氣象局資料範例	18
表 3.3 資料分類	30
表 4.1 範例租借站借還資料	32
表 4.2 時間點切割範例	35
表 4.3 範例 A 站起訖點移動機率	38
表 4.4 範例 A 站租借 10 台車的訖點車數估算	38
表 5.1 平日最適車數計算範例—台北 YOUBIKE 捷運市政府站	43
表 5.2 決策樹數量的 RMSE 值比較	45
表 5.3 隨機森林特徵值	45
表 5.4 降水量分級	45
表 5.5 四組特徵值組合	46
表 5.6 四種特徵值與各站點 RMSE 測試結果	47
表 5.7 模擬測試範例結果	53
表 A-1 圖 5.8 數值	68
表 A-2 圖 5.9 數值	69
表 A-3 圖 5.10 數值	70
表 A-4 圖 5.11 數值	71

## 圖目錄

圖 1.1 全球公共自行車共享系統分佈 .....	1
圖 2.1 視覺化圖表(BARGA ET AL., 2014).....	12
圖 2.2 時間區別方式(MONTOLIU, 2012).....	13
圖 2.3 自行車使用特性圖(FROEHLICH ET AL., 2009) .....	13
圖 3.1 2014 年 1 至 10 月 YOUBIKE 車站數變化圖 .....	18
圖 3.2 2014 年 1 月台北 YOUBIKE 平日與假日借車趨勢圖 .....	19
圖 3.3 2014 年 1 月台北 YOUBIKE 平日與假日還車趨勢圖 .....	19
圖 3.4 台北 YOUBIKE 捷運市政府站星期一借車數量比較圖 .....	20
圖 3.5 台北 YOUBIKE 2014 年 1 月公館站起訖點統計圖 .....	21
圖 3.6 每 5 分鐘結算一次 .....	21
圖 3.7 每 30 分鐘結算一次 .....	22
圖 3.8 每 360 分鐘結算一次 .....	22
圖 3.9 視覺化工具操作流程 .....	23
圖 3.10 地理範例資料 .....	24
圖 3.11 台北 YOUBIKE 系統視覺化範例截圖 .....	25
圖 3.12 台北 YOUBIKE 站點熱點圖截圖 .....	25
圖 3.13 台北 YOUBIKE 站點資訊截圖 .....	26
圖 3.14 視覺化工具地圖圖標截圖 .....	26
圖 3.15 台北 YOUBIKE 捷運市政府站站點資訊圖 .....	27
圖 3.16 台北 YOUBIKE 捷運市政府站起訖關係資訊圖 .....	28
圖 3.17 台北 YOUBIKE 捷運市政府站的起訖旅次總數圖 .....	29
圖 4.1 群眾運補研究方法流程圖 .....	31
圖 4.2 時段定義 .....	32
圖 4.3 範例租借站 A 之借還資料示意圖 .....	32
圖 4.4 台北 YOUBIKE 2014 年 1 月國興青年路口站訖點總數 .....	33
圖 4.5 已借出但尚未歸還之車流量示意圖 .....	38

圖 4.6 起訖點移動示意圖 .....	38
圖 5.2 隨機森林決策樹數量與誤差比較圖 .....	44
圖 5.3 特徵值重要性比較圖 .....	46
圖 5.6 台北 YOUBIKE 以 K-MEANS 分成 10 群分群圖 .....	51
圖 5.7 運補車運補策略示意圖 .....	51
圖 5.8 群眾運補與運補車失敗率比較 .....	54
圖 5.9 群眾運補是否預測失敗率比較 .....	54
圖 5.10 運補車數量與群眾運補失敗率比較 .....	55
圖 5.12 群眾運補與運補車運補經過 100 次隨機選取資料測試後的失敗率比較	57



# 第一章 緒論

## 1.1 研究背景

公共自行車共享系統（ Public Bike Sharing System，PBSS ）從 1960 年代開始發展，至今全球已經有超過三百個自行車共享系統在營運，PBSS 被認為是用來解決交通「最後一哩路」的方法，加上近幾年環保意識的抬頭，騎乘自行車能有效降低噪音及空氣汙染，此外人民也越來越重視健康及休閒生活，騎乘腳踏車已成為閒暇時民眾的休閒選項之一。如圖 1.1 所示，目前全球已有超過 375 個 PBSS，台灣也跟上此波風潮，台灣最早營運的 PBSS 的為高雄市的 City Bike，而現在全台灣則有七個正在營運的 PBSS，其中台北 YouBike 的站點數及使用率不僅是全台居冠，週轉率更曾達世界第一（最高曾每台車單日平均被租借 12-13 次）。



圖 1.1 全球公共自行車共享系統分佈

最早在台灣營運的高雄 City Bike( 簡稱 CityBike )截至 2016 年 8 月為止共有 184 站，目前實行前一小時免費，因此部分民眾可以將其作為長時間移動的交通工具。而台北 YouBike 主要由捷安特巨大機械股份有限公司建置和營運，在最初推行時，以前 30 分鐘免費吸引民眾騎乘，經過了多年，已經有許多民眾習慣騎乘 YouBike，至 2015

年 4 月已取消前 30 分鐘優惠。新北市和台中市也是近年巨大公司極力推廣 YouBike 的城市，新北與台北的相鄰也使得雙北的整體自行車範圍擴大。

共享自行車的普及帶來民眾交通上的便利，其成敗則取決於是否能成功地處理「租借站分布狀況」、「租借站車數控管」及「民眾使用慣性」等三個議題。其中，「租借站分布狀況」為一選址問題，包括如何選擇最佳租借站位址及停車柱個數等；「租借站車數控管」為租借站每日各時刻最佳的車輛存貨管理問題，包括如何訂定各時刻(或期間)的最佳自行車數(亦即空位數)、如何在租借站間運補自行車以達成其最佳自行車數等；至於「民眾使用慣性」為深入分析民眾使用習慣，藉此可協助預測租借行為對自行車庫存量的影響。

## 1.2 研究動機

目前大部分的 PBSS 已進化到第四代(DeMiao, 2009)，其特點為無人化自動管理租借站(Kiosk)。Kiosk 系統的優點為簡易的操作以及 24 小時營運的特性，使得民眾在使用上相當方便，不過此類系統通常不允許借出之自行車隨意亂停放，必須停在系統的空車柱才能成功還車。PBSS 最需要處理的服務問題為「無車可借」和「無位可還」，前者為當使用者到達租借站欲租借時卻發現所有自行車皆被借走；後者則為當使用者到達租借站欲還車卻發現沒有空位可以還車(亦即全部車柱皆停滿)。雖說「無車可借」和「無位可還」皆會造成使用者困擾，但兩者相較又以後者影響較大，這是因為使用者在「無車可借」時尚可選擇搭車或走路等其它交通方式；而在「無位可還」時，如何趕緊找到附近租借站的空車柱還車即為使用者的唯一要務。對營運商而言，處理「無位可還」的「移車」與「無車可借」的「補車」其實是互補的作業，目前最常見的作法大都使用卡車自某租借站「移車」至另一租借站「補車」，然而此作法通常緩不濟急、運補效益不彰。以台北 YouBike 來說，因都會區內交通較為繁忙，加上租借站點所跨越的距離較遠，受到卡車移動範圍和數量上的限制，少量的卡車通常難以滿足大量租借站的同時缺車、缺位運補需求，若要達到每個租借站皆能迅速被運補

的理想情境，只能動用大量的卡車和人力做運補，而這又將耗費大量的人力和財力成本。以下我們以 YouBike 的資料粗估利用卡車的運補效益。

如表 1.1 所示，從此參數可推得，平均一分鐘可以卸載一台車所以每 20 分鐘可以卸載 20 台車。而站與站之間的移動距離為 10 分鐘，所以每 30 分鐘可以卸載 20 台車，則每小時可以卸載 40 台車。

表 1.1 以卡車運補之粗估參數範例

參數名稱	數值	備註
車站數	200	
卡車數	10	
自行車數	5000	
自行車周轉率	8	一台自行車一日可被借幾次
系統營運時間	18	單位：小時
卡車於站點間移動時間	10	單位：分鐘
卡車卸載一台自行車時間	1	單位：分鐘

接著估算平均每站每小時的租借數：

假設一台自行車平均一天可以被租借 8 次，則 5000 台自行車在工作的 18 小時內，總共被租借 40000 次，因此平均每站每小時租借數為 11 次，即平均每  $1/11=0.09$  小時就會被租借一次。

綜合上述參數而言，10 台卡車每小時可以運補 400 台自行車，而整個系統有 200 個站點，平均只要每個站點連續 2 次( $400 \div 200$ )的租借即可抵消卡車運補 400 台自行車的效益。而租借兩次自行車的時間為 0.18 小時；換句話說，10 台卡車在一小時內所作的努力，只要花費 0.18 小時就會被抵消，如果說希望每次的運補能夠維持系統一小時的平衡，則須將卡車增加至 56 台，不只需要更多的卡車與相關人力花費，這些卡車進而造成的空污、交通壅塞等環境成本更與原先 PBSS 的設置宗旨背道而馳。

從上述粗估分析可知道，目前僅靠少數卡車要平衡運補需求，勢必效益其差，因此本研究提出「群眾運補」(Crowdsourced Repositioning)的概念，有別於傳統利用卡車與人力運補的方式，主動招募志願群眾接受運補任務，群眾依系統指示先至指定的多車租借站租車，再騎至指定的少車租借站還車。在給予適當獎勵的情況下，只要有民眾願意接受號召志願騎乘運補，則可以同時在多個租借站進行運補，相較於數量和移動方式較受限制的卡車運補，群眾運補更便宜、有彈性、且有更好的效益。

此外，過去相關的 PBSS 營運研究大都將一天切割成數個時段，假設每時段各租借站的租借需求為已知的固定常數（通常取真實資料之平均值），並以此假設的靜態數值為基礎來推導系統建置或自行車運補之數學規劃模式。然而現實的租借需求常會依不同租借站、日期、時刻、天氣、突發事件等因素而可能有不小幅度的變化。數學上來說，也就是這些平均之租借需求其實蠻常伴隨著較大的變異數，因此若將租借需求視為單純的靜態平均值，可能會讓規劃出來的結果與現實差距太大。本研究亦以政府公開資料(Open Data)為基礎，先以機器學習方式深入探討租借需求的變動趨勢，並藉此來預測短時間內預期的租借需求變化，再修正數學規劃模式。

### 1.3 研究目的

針對群眾運補的概念，本研究以台北 YouBike 從 2014 年 1 月至 10 月的所有交易資料為資料基礎，並利用兩個數學模式去定義群眾運補的運補方式，計算群眾運補的結果，而為了提高運補的準確度，利用隨機森林演算法預測未來的可能借還車需求，最後提出一個結合最適車數模型以及預測模型的群眾運補模擬流程。研究目的主要分成以下主題：

#### 1. 資料分析

本研究所採用的資料集為台北 YouBike 的真實交易資料，會分析出台北 YouBike 的交易特性，並擷取出所需要的資料用於第四章的數學模式及預測模型。

## 2. 自行車租借站最適車數數學模式

廣設租借站為 PBSS 成功要素之一，而每個租借站的特性、容量皆不同，若放置太多則容易滿車，而放太少又容易缺車，因此每站何時該放幾台車是 PBSS 極為重要的營運決策問題。針對此問題，本研究擬先以歷史租借記錄平均值為基礎，假設所有的租借需求完全可藉運補（假設不計成本）來完全滿足的理想情況，提出一個數學規劃模式用以計算符合此理想情況之各站各時刻的理論最適車數，而該最適車數值即可用於之後因應需求變動的即時運補決策參考，藉此為目標來推算適宜之群眾運補策略。

## 3. 隨機森林預測模型

為了因應需求的高度變動性，必須對租借資料有更精準的預測方式，本研究採用 Breiman (2001) 提出的隨機森林模型作為預測模型，利用台北 YouBike 2014 年 1 月至 10 月的交易資料建構出各站點的借車、還車預測模型。

## 4. 群眾運補策略

群眾運補需要在適當的時機召集適當數量的志願群眾幫忙騎乘某些起訖點（Origin-Destination, OD）組合的行程，因此必須要先掌握每個租借站各個時段的需求變動趨勢（例如在捷運站附近可能使用量較為不固定、公園或是學校的需求尖峰多是落在早上下午運動及上下學時段等），在此我們以 30 分鐘為一期，藉由分析歷史租借資料，可以了解各個站點在不同期的使用者習性，本研究提出之群眾運補策略簡述如下：首先建立橫跨兩時段（當下及下一期）之數學規劃模式，擷取當下即時租借資料（各站車數、前期已借出但尚未還之車數等），由歷史租借資料推估當下這一期間各站預期發生的還車人次，再由隨機森林模型來預測即將在這一期發生的租車人次，以儘可能達成下一期各站之最適當車數為目標，建構一個數學規劃模型以計算當期各站應該召集多少人去騎乘哪些 OD 行程，每期重複此流程，直至當日結束為止。

## 5. 群眾運補模擬系統

為了廣泛測試群眾運補效果，本研究實作了一個模擬系統，以分析過的租借資料為基礎來模擬產生租借需求，並結合先前的最適車數數學模式以及隨機森林預測模型，來模擬群眾運補決策。其模擬流程盡可能接近真實公共自行車共享系統的營運流程，先模擬產生一整天的租車需求資料（屆時可否成功租還將視當下車站庫存量而定），在各期開始時，依當下各站的自行車庫存等資料利用隨機森林預測模型來預測即將發生的租借需求，再預估當下已上路被租借的自行車之還車時刻與租借站，以事先計算出來的下一時段之最適車數數值為運補目標，求解橫跨兩時段（當下及下一期）之群眾運補數學規劃模式，依其最佳解下達當下的運補指令，假設所有運補指令皆可被成功執行（指派成功率亦可自行調整），再讓模擬系統依其原先產生之租、還車事件照表進行，直至該期結束（亦即下一期開始）；重複每時段的模擬直至一整天的交易結束，即可總整一整天的模擬效益。

### 1.4 論文架構

本論文之架構如下：第二章為文獻探討，先介紹過去學者針對 PBSS 的相關研究（例如在特定時間各站的最適當車數該如何給定，或是從實際資料中挖掘出可使用的相關特性，並加以將這些特性套用在數學模式或是實際生活層面做應用）。第三章為資料分析，以台北 YouBike 2014 年 1 月至 10 月的資料為分析對象，從中找出相關的特性，並將這些特性納入第四章的數學模式中。第四章為群眾運補數學模式，本研究會提出最適車數以及群眾運補的數學規劃模式，其中最適車數之數學模式旨在計算理論上各站各時段的最佳自行車庫存量，而群眾運補之數學規劃模式有賴更精準的需求預測模型（在此我們使用「隨機森林」來協助預測租借需求），接著提出兩時段（本期與下一期）之群眾運補的數學規劃模式。第五章將以第三、四章的資料分析結果與數學模型作為基礎，提出一個群眾運補的運補模擬流程，以測試群眾運補之效益，並

以台北 YouBike 的實際交易資料作為資料依據，進行數值分析與討論。第六章總結本研究，歸納本研究之貢獻，並提出未來可行之研究方向建議。



## 第二章 文獻回顧

本章將介紹先前的 PBSS 相關文獻，2.1 節會介紹靜態運補與動態運補的相關文獻，討論如何透過運補方式達到各站的平衡。2.2 節會介紹利用資料分析的方式了解使用者的使用趨勢，以及預測使用者需求的相關研究。2.3 節會介紹採用不同營運模式，是否能改善目前公共自行車共享系統的研究。2.4 節介紹本研究採用的預測模型，隨機森林演算法。最後 2.5 節為此文獻探討的小結。

### 2.1 運補策略之相關文獻

在 PBSS 營運期間，各站的自行車數隨時都在變動，起訖點間的自行車移動使得租借站容易發生車數不平衡的情況，若空車或空位過少，則會降低使用者使用的滿意度，因此如何維持各站車數的平衡極為重要。目前在自行車運補的文獻中，主要可以分為「靜態自行車運補」(Static Bike Repositioning)及「動態自行車運補」(Dynamic Bike Repositioning)兩大類問題，以下將分別描述此兩者之差異及相關研究。

#### 2.1.1 靜態運補

靜態運補是指在整個系統沒有任何交易(如深夜期間)的情況下，執行站間的自行車運補作業。此類運補問題通常會先設定隔天系統營運時各站的初始車數，以能達成該初始數為主要目標，使用卡車執行運補。研究的議題是卡車之路徑規劃、最小運補成本及運補工作量的分配等。

廖敏婷(2012) 加入運補車行走距離、工作量與運補車間工作量的差異做考量，以求出滿足各租借站缺車、缺位的狀態。而隨著站點與運補車數量的增加，求解時間也隨之上升，因此作者除使用整數規劃模式外，亦提出粒子群演算法加速求解。Raviv et al. (2013) 以靜態運補為基礎，提出兩個整數規劃模式，在不同路徑規劃及決策模式下作靜態運補，最後利用真實資料做測試，他們提出的模型可以在合理的時間內求出品質不錯之解。

## 2.1.2 動態運補

靜態運補僅適用於無或少人使用的期間（譬如夜間）進行，而平常使用者較多時的車輛調度方式即為動態運補。因為租借站的車輛庫存量會受到動態的使用者需求（e.g., 民眾同時借還車）而振盪，這也嚴重地影響了站間的車輛調度決策，因此動態運補較靜態運補更重要、同時也更難做。本研究亦探討動態運補，但與過去有兩大不同點：(1)使用即時資料做當下的運補調度參考，有別於文獻大多採用靜態的平均租借資料；(2)以號召群眾的方式而非如傳統使用卡車的方式來運補。以下回顧動態運補的相關文獻。

張立蓁(2010) 提出一個兩階段模式求解，先假設運補用的卡車無數量上限，第一階段以整數規劃模式預估在相鄰兩期間各起訖站間各需要多少運補車，第二階段再試圖指派有限的卡車在站間運補以極小化未被滿足的運補需求；而洪菁蓬(2012) 以張立蓁的模型為基礎，依據不同程度的租借資訊，提出三種不同情境：(1)僅知起訖點，無法確知前往哪些租借站租還；(2)已知起訖租借站，但未知各租借站之歷史租借需求變化趨勢；(3)已知起訖租借站，且將各租借站之歷史租借需求變化趨勢列入考慮，分別以最小成本多元商品網路流量模式求解動態運補問題，該研究證實若同時將歷史租借資訊以及即時資料列入考量，的確能有效地改善公共自行車共享系統的效益。不過，此兩研究顯示整數規劃模式僅能處理小規模（e.g., 11 站、8 期、2 卡車，洪菁蓬, 2012 )的運補問題，而即使使用諸如粒子群演算法( Particle Swarm Optimization, PSO )等啟發式演算法，求解效果其實還是不佳（與上述 11 站小例子相較，誤差達 10% ）。

Contardo et al. (2012) 提出一個整數規劃模式，其求解目標為最小化使用者租還不滿足的情況，他們的研究著重在求解的速度，利用兩種不同的分解法以及演算法來幫助求解，尤其針對中大型的系統，他們的方法能夠在很快的時間就找到上下界，但是仍然有最佳解差距過大的情形。

Vogel et al. (2014) 提出一個混整數規劃模式，在他們的模式裡能同時解決自行車配置以及最小化民眾需求，但由於他們考量的是整天的動態配置情況，整體的變數會

因為時間及站數的增加使得求解速度變慢，因此他們使用啟發式演算法及最大鄰近搜尋法協助求解。

有別於大部分文獻使用卡車的運補方式，廖敏婷（2012）首度提出暫時人力配置方式，假設可以招募的暫時人力有其上限，以整數規劃模式求解出最佳的暫時人力以在某些時段於某些滿車之租借站協助處理還車需求，並假設可找到暫時人力來承接某些起訖組合的運補需求，該研究仍以各站各時段的歷史租借記錄平均值為基礎來設計其數學模式，而本研究亦將以其模式為基礎，再進一步考量現實的租借資料變動性。

周佑賢（2015）更進一步依據歷史租借記錄的分析結果，以兩階段式車輛途徑的方式處理動態運補，第一階段先劃分責任區域，第二階段再分別對各區域求解一台運補車之最佳運補路徑與卸載計劃。其第一階段用 K-means 分群演算法將所有車站分群，以 30 分鐘為單位將每日分為 48 期，再以淨流量（各期各站的還車數減去借車數）及站間距離為主要考量因素做分群。接著在求解第二階段的各區域單一運補車運補計劃時，提出「運補相斥性」的性質，即運補車在做運補路徑規劃時，不會傾向運補有連續兩期同為缺車或是缺位的站點，因為運補車在缺車站點已經將車上的自行車都卸下，之後應該到缺位站點將自行車放置運補車，加入此限制有助於運補車規劃路徑。此外，該研究首度將 YouBike 於每租借站設置「臨停區」的現實作法列入數學模式的設計當中，一併求解臨停區與租借站間的自行車調派方式。

總結來說，目前為止的動態運補相關文獻大多將一日切割成數個時段（通常以 30 分鐘為一期），假設各站各時段的租借交易資料（起訖時地）皆為已知，再試圖規劃卡車在各期間的最佳移動方式以及各期內的最佳卸載車數。解法大致分成整數規劃模式或啟發式演算法，前者雖然理論上可求得最佳解，但實際上常僅能處理規模較小的問題，實用性不佳；後者（e.g., PSO）雖可處理大規模運補問題，但其實效果也不佳。此外，上述文獻皆以靜態租借記錄平均數值來規劃，完全無視當下可能與歷史的租借平均值差異甚大之現實（real-time）租借資料。

本研究擬以現實租借資料為基礎，提出當期群眾運補數學模式，使用本期的即時資料以及過去的歷史資料，預估本期即將發生的租還車趨勢，決定最佳的群眾運補方式，以達到下一期期初（本期期末）理想的車輛庫存為目標。這樣的作法因為僅考慮單期，因此可很快求解，而在下期時因為可再依當下最新的租借資料來調整參數，因此也更符合現實運補的需要。然而，本作法有賴對租借資料的變動趨勢有深度的分析與理解，因此亦必須在資料分析部分加以著墨。

## 2.2 資料分析之相關文獻

資料分析相關文獻中，主要探討如何挖掘各種會影響民眾騎乘自行車的因素，其中有探討使用者行為、天氣因素、時間、距離及站點位置等因素，此類研究會從挖掘資料的結果衍生出相關的議題，例如視覺化工具、需求預測模型。

有一些研究是以特定地區為分析對象，Gebhart and Noland (2014) 以華盛頓地區的資料配合氣候資訊去研究當氣候改變時，自行車使用率的變化。該研究發現關於下雨天不會有人騎自行車的觀點是錯誤的，對於沒註冊的使用者來說，下雨天偏向不騎車，但是有註冊的使用者在下雨天還是會騎腳踏車；此外他們也發現有私人腳踏車的人會傾向在惡劣的氣候騎乘公共自行車，以避免自用腳踏車受損。Barga et al. (2014) 以波士頓、華盛頓和芝加哥的資料為主，做出一個互動式的視覺化系統以用來觀察資料特性，如圖 2.1 所示，此視覺化工具可以依據時間、地點調整參數，顯示不同的站點特性或是分群結果，最後將分析後的資料做分團問題(clique problem)，從中找出最大的一團進而做分群，並能得知在任何時刻的最大自行車交易流量的所在位置。O'brien et al. (2014) 分析 30 個不同地區的公共自行車站點，列出這 30 個公共自行車站點不同之處，一一分析各系統的站數、交易波動、天氣狀況以及星期的不同。Sarkar et al. (2015) 分析了 10 個不同的公共自行車站點，與 O'brien et al. (2014) 不同之處在於他有將資料做正規化處理，例如各系統的站點數、站點容量皆不同，數量不同則難

以用「數量」表示其缺車缺位的危及程度，因此其將這類數據轉換成百分比做比較，並依據這些特性，將站點做分群分類。

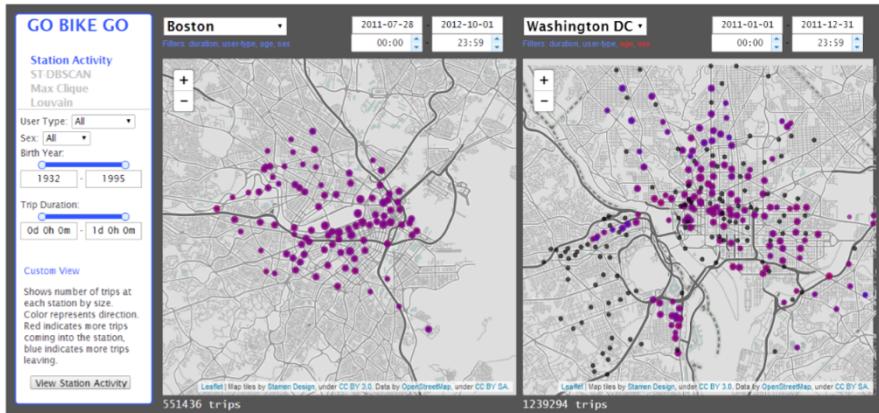


圖 2.1 視覺化圖表(Barga et al., 2014)

另有一些研究是從已知資料延伸出相關方法，O'Mahony (2015) 以資料分析的結果放入最佳化模型以解決卡車在隔夜運補的路徑問題，以及在尖峰時段用量的問題。Vogel et al. (2011)利用資料探勘的方法尋找公共自行車交易的波動，並利用地理資訊系統的方法，找出最適當的設站位址，以減緩站間的運補需求。

Montoliu (2012) 提出利用 Latent Dirichlet allocation (LDA) 方法找尋使用者移動時的特性，想要找出是否有類似符合早上多借、下午多還的租借站等例子。該研究發現在中型城市的使用者常有使用特性，其提出的方式可以抓取在特定期間、特定站的使用者行為，其時間切割的方法如圖 2.2 所示，將一天切為八個時段：7:30~9:00AM、9-11AM、11AM-1PM、1-3PM、3-5PM、5-7PM、7-9PM、9-10:30PM。由於絕對的數字會增加判定趨勢的難度，因此該研究設定一個門檻值  $\tau$ ，在該時段的交易淨流量若為正值且大於  $\tau$  的話，則給予  $\nearrow$  增加趨勢的符號；該淨流量若為負值且其絕對值大於  $\tau$  的話，則給予  $\searrow$  下降趨勢的符號；若都沒增加或是下降，則給予  $\leftrightarrow$  持平的趨勢。此粗估方式的優點在於不會受到偶爾突發的租借影響太大。

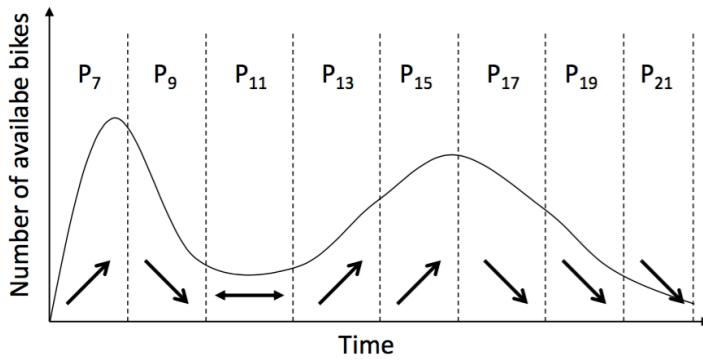


圖 2.2 時間區別方式(Montoliu, 2012)

在預測需求的相關文獻中，Froehlich et al. (2009) 以巴塞隆納的自行車共享系統為資料來源，分析出以下特性：

(1) 歸納出使用者的特性

使用者特性因為各站容量不同，交易波動也不同，為了避免受到交易「數」的影響，此文獻將數字正規化後再去比較，得知個別站點或是系統整體需求在平日與假日都呈現不同的趨勢。

(2) 歸納出自行車使用特性與地理位置的關係

自行車使用特性與地理位置的關係如圖 2.3 所示，其作法為將交易數量轉化成「活動分數」，利用動態時間扭曲演算法(Dynamic Time Warping)找尋相似波動的曲線並將之分類，發現有相似行為的站點會聚集在一起。

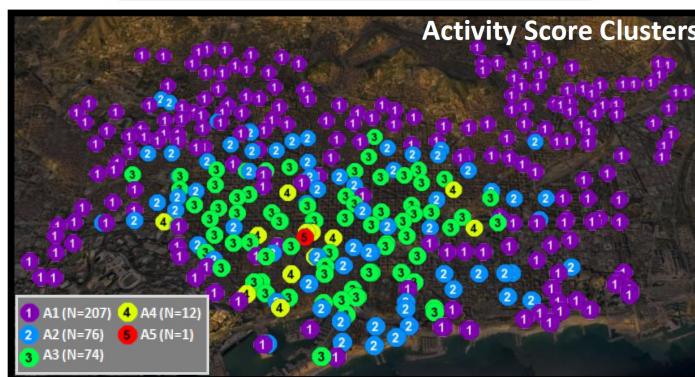


圖 2.3 自行車使用特性圖(Froehlich et al., 2009)

(3) 預測未來使用者需求並歸納出主要影響需求的參數

使用者需求預測模型中考量的參數有當下時間、當下的車數、要預測的下個時間範圍，並利用四個不同的方法做預測，其預測誤差大約在 15%左右。

預測相關文獻中, Kaltenbrunner et al. (2010) 以時間序列預測中的 Auto-Regressive Moving Average(ARMA)作為其預測模型，由於 ARMA 必須先將資料做過前處理，否則產出的預測會較不準確，作者認為其可再跟其它演算法比較合適性。Yoon et al. (2012)以都柏林的腳踏車資料為主，發展一套基於 ARIMA 模型的預測模型，其改善模型有加上季節變化和空間地理位置相關的因素，他認為過去的研究採用貝氏分析或是舊的 ARIMA 模型都仍有缺陷，而該研究提出來的模型預測準確度皆優於過去提出來的方法。Rixey et al. (2013)分析了三個美國的公共自行車，他們利用多變量回歸分析方法產生預測模型，並列出會影響模型的多個因素。Cagliero et al. (2016)以紐約城市自行車共享系統為主要分析對象，提出了一個自行車車位預測系統(STation Occupancy Predictor)，將每天的新資料匯入系統中，並針對當下車數做出短期以及中期的車數預測。該系統採用 AODEsr 和簡單貝氏等兩種貝氏分類器，作者認為採用分類法作為預測模型比採用回歸類型的演算法來的好。Yang et al. (2016)以中國杭州的自行車共享系統資料為基礎，發展出可以預測借、還車的模型。對於還車模型作者的概念為還車量取決於已被借出的車量，利用此想法建構一個演算法求出預測還車量。此模型分析過去的歷史資料，列出各站點之間的移動機率，利用借車的數量乘以此移動機率算出未來各站點的還車值。而為了加速預測的計算速度，會依據歷史資料求出機率極低的借還車站點組合(即車程太遠的站點組合)，在計算預測還車時則完全不會考慮這些機率太低的可能訖站。而借車模型部分則利用隨機森林演算法作為預測模型，該方法利用歷史資料擷取出對於杭州自行車共享系統重要的特徵值，例如星期、天氣、氣溫等，並依據這些特徵值建立一個可以預測各站點使用需求的隨機森林模型，經由與其它預測演算法相比後，隨機森林演算法的預測結果是最佳的。

在過去預測的研究中，因為各站點容量皆不同，需要將借車或還車輛正規化或是做一些轉換，但是哪種轉換是比較合理的，目前也沒有文獻給予肯定的答案。而有多學者採用不同的預測模型，如線性回歸、時間序列預測、隨機森林模型，目前 Yang

et al. (2016)在其所作的測試中發現，隨機森林模型的表現比其它常見的預測模型來的佳，本研究也將以隨機森林模型作為主要的預測模型。

### 2.3 營運策略相關文獻

目前第四代的公共自行車共享系統標榜使用者可以自行租還自行車的營運模式，但此作法在某些時地容易衍生無車可借、無位可還的情形，因此有部份研究探討如何透過營運方式的改變，以解決這些問題。

廖敏婷 (2012) 提出兩種暫時人力的配置方式：其一為派遣人員在某時刻至某些租借站協助無位可還的民眾還車(譬如滿車時，可逕行將自行車歸還給該人員即可)，其二則是透過調降費率、結合商品行銷等手法，讓使用者願意騎乘至自行車較少的租借站協助運補。該研究假設一整天各站各期的需求量為其歷史租借記錄的平均定值，以此計算出一整天各站各期的最佳運補起訖點組合（此即為欲號召之志願騎乘者行程），其缺點為忽略了現實租借資料的變動性。

Kaspi et al. (2014) 提出了針對共享車輛系統的預約機制，可適用於公共自行車和公共汽車共享系統。該研究提出了提前預約模式，即在租借前就先需告知要在哪個站點還車，如此可以及早知道使用者什麼時後會來車站租還車。該研究並建構使用者行為模型，探討使用者從到租借站至離開之間所作的決策有哪些，接著利用系統模擬的方式去估算總使用時間，發現任何預約方式的使用時間都比沒有預約來的短，因此認為採用預約策略對於共享車輛系統的營運是有幫助的。

### 2.4 隨機森林演算法

為了預測下一階段會有多少人來租借，本研究擬採用隨機森林演算法來實作。隨機森林演算法是 Breiman (2001) 提出來的機器學習演算法，隨機森林是一個包含多個決策樹的分類器，由多個決策樹決定輸出的類別而定，也因其是多個決策樹，因此則稱為森林。隨機森林演算法的演算法流程為：

假設現在有  $N$  筆資料，各有  $M$  個特徵值：

1. 輸入特徵值數目  $m$ ，用來決定決策樹上節點的決策結果（其中  $m < M$ ）
2. 從  $N$  個訓練樣本中以有放回抽樣的方式取樣  $N$  次，將這些樣本集結為一個訓練資料，而為抽到的則作預測，用以評估誤差。
3. 對於每一個節點隨機選擇  $m$  個特徵值，依據這  $m$  個特徵值決定決策樹上的節點。

隨機森林演算法因為是以多個決策樹決定分類標準，因此有很高的分類準確度，在演算法的運作效果上，其訓練和預測的速度很快，適合用在本研究的即時資料。隨機森林亦會提供每個特徵值的重要性評估，可提供在做特徵值驗證時做參考。

在過去公共自行車的文獻中，Yang et al. (2016)將隨機森林演算法用在借車的預測模型，其採用的是 Breiman (2001)提出來的版本，他們發現隨機森林的表現比其它常見的預測方法來的好。本研究針對預測模型也同時測試了「線性回歸」以及「時間序列方法」，結果也是以隨機森林的表現較佳，與 Yang et al. (2016)不同之處在於由於選用資料集的不同，所選取的特徵值以及產生的決策樹數量不同。

## 2.5 小結

PBSS 在近幾年才興起，目前相關的文獻大致可分成租借場址選擇、自行車運補方式、以及租借資料分析方式等三方面，其中後兩項與本研究較相關。運補部分又以「動態運補」更為重要，大多文獻使用整數規劃或啟發式演算法求解最佳的運補卡車移動方式以及卸載計劃，然而這些方式有兩大缺點：(1)卡車之運補效益不佳；(2)使用的參數多為租借平均值，未考量現實租借需求的大變異性，可能導致規劃結果與現實差距甚大。在租借資料分析的相關文獻部分，因為 Gebhart and Noland (2014)提出的視覺化圖表可以幫助系統營運商對各系統間的狀況一目瞭然，本研究擬參考其視覺化方式實作租借資料之視覺化工具。

### 第三章 資料分析與視覺化工具

為了讓數學規劃模式更貼近實際，首先必須先對租借資料有更進一步的了解，本研究以台北 YouBike 從 2014 年 1 月至 10 月的所有交易資料作為分析對象，因為台北 YouBike 的週轉率曾為全球最高（單日每車平均被租借 12-13 次）且至今尚未有相關文獻對其租借資料做系統性的分析。3.1 節介紹台北 YouBike 的資料型態。3.2 節介紹欲了解的資料特性，並將這些特性利用視覺圖表方式呈現。3.3 節為本章小結。

#### 3.1 資料簡介

台北 YouBike 提供的資料格式如表 3.1 所示，每筆資料都包含其借車場站、還車場站、借車時刻、還車時刻。從 2014 年 1 月至 10 月，平均每個月的借還資料共有 150 萬筆，十個月總資料量則約有 1500 萬筆，而車站數變化如圖 3.1 所示，這 10 個月大約增加了 20 個站，本研究將以這 1500 萬筆資料為資料來源，分析出所需的資料，提供之後的數學模式使用。

表 3.1 租借資料範例

借車時刻	借車場站	還車時刻	還車場站
2014-01-01 22:58:40	興雅國中	2014-01-01 23:02:32	捷運市政府站(3 號出口)
2014-01-01 18:30:26	龍門廣場	2014-01-01 19:13:43	捷運市政府站(3 號出口)

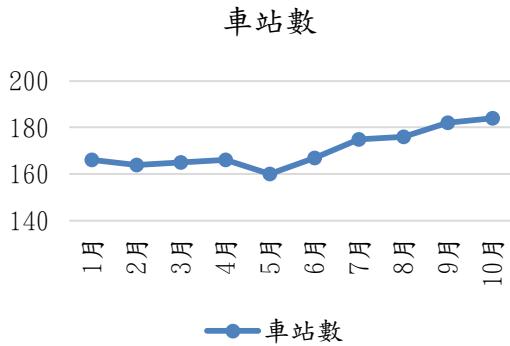


圖 3.1 2014 年 1 至 10 月 YouBike 車站數變化圖

除了交易資料外，當時的氣溫、下雨時間也會影響腳踏車的借還情況，因此本研究抓取中央氣象局 2014 年的氣候資料，如表 3.2 所示，氣象資料為每小時記錄一次，本研究將僅採用降水量與降水時數等兩項資料。

表 3.2 中央氣象局資料範例

觀測時刻	氣溫(°C)	相對溼度	風速(m/s)	降水量(mm)	降水時數(hr)
1	14.6	60	1.0	1.5	0.2
2	15.2	67	0.8	0.0	0.0

## 3.2 資料分析

針對資料分析可以分成「平日與假日交易趨勢」、「晴天與雨天」、「起訖點趨勢」、「時間切割」等四部分。

### 3.2.1 平日與假日交易趨勢

如果民眾使用公共自行車的目的為交通、休閒娛樂，那直觀而言，平日與假日的交易趨勢應不同。針對此趨勢，本研究將每日的借車、還車記錄繪製在圖表上，如圖 3.2、圖 3.3 所示，分別代表 2014 年 1 月台北 YouBike 的平日與假日借車、還車交易量，以每五分鐘結算一次所有站的總交易量，可以發現平日與假日之交易趨勢很不一

樣，譬如平日借還車的高峰點皆大概在 7:30~9:00 和 17:00~19:00 之間；而假日的高峰則大概在下午大約 14:30~18:00 期間。在平日時或許可解讀為有許多民眾將自行車當作通勤的交通工具使用。此外在圖 3.2 中有一個箭頭指向的趨勢，該趨勢的時間點是 1 月 1 日元旦連假的期間，可以發現該日與其它日的趨勢完全不同，所以可以從圖中得知 YouBike 的交易會受平日、假日以及連假影響。

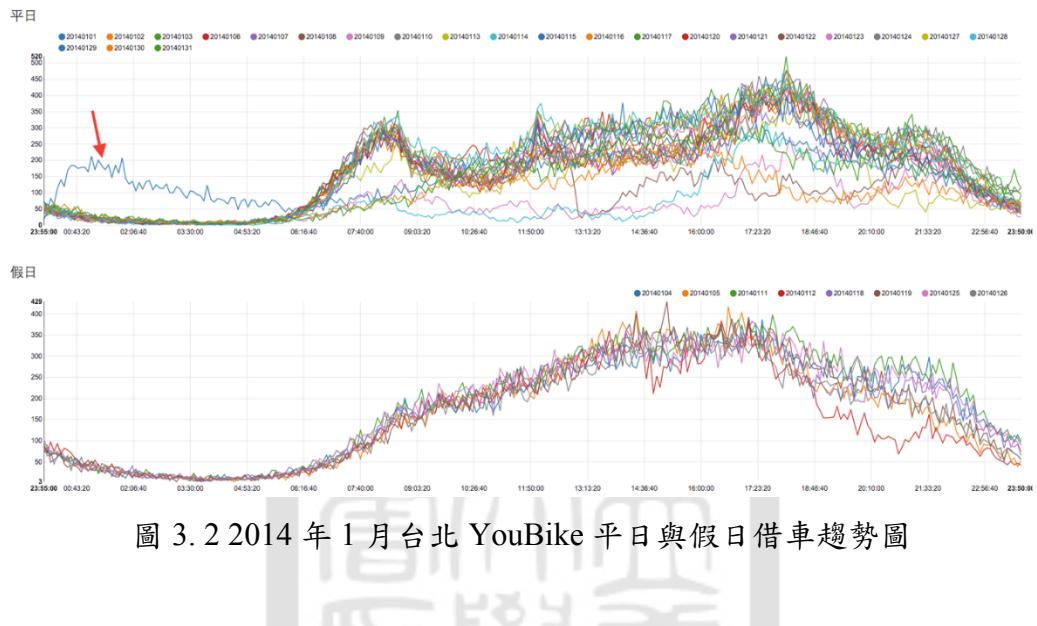


圖 3.2 2014 年 1 月台北 YouBike 平日與假日借車趨勢圖

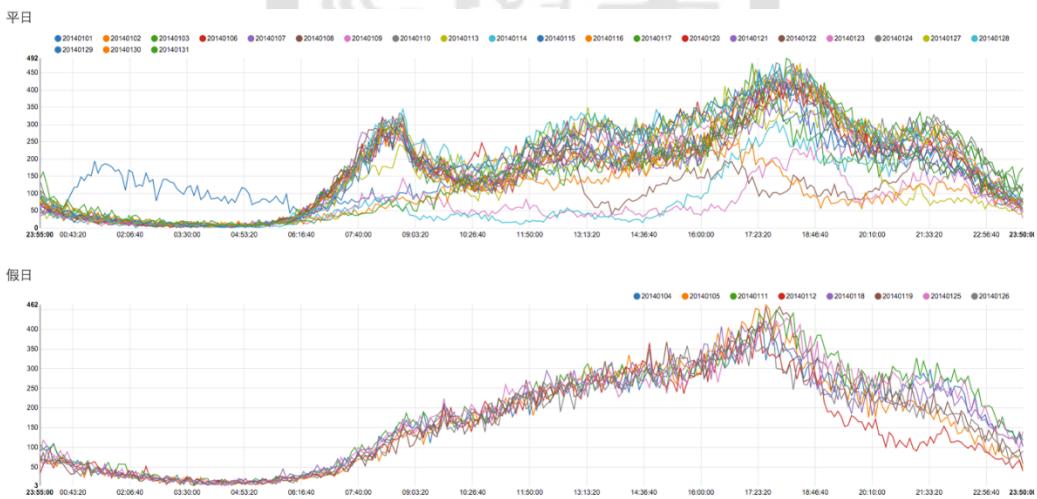


圖 3.3 2014 年 1 月台北 YouBike 平日與假日還車趨勢圖

### 3.2.2 晴天與雨天

Gebhart and Noland (2014) 發現以華盛頓地區來說，下雨天仍舊有民眾會願意騎乘自行車。為了瞭解雨天是否會影響台北 YouBike 的交易數量，本研究將 2014 年 1 月至 5 月的各站不同星期、不同時間區間的所有交易量算出並繪製在圖表上，以圖 3.4 捷運市政府站星期一借車數量為例，圖中每一個點表示的是星期一中每半小時的交易數量，其中菱形(◆)表示的是該時段有下雨的交易數量。從中可以看出下雨時的借車數量都普遍偏低，但借車數還是有高有低，也就是雨天的借車數會受到當時雨量的大小所影響，針對此部分可以利用中央氣象局提供的「降水量」計算。從分析結果可以歸結出雨量的多寡會影響民眾借車的意願，在第四章則可以將降水量放入模型當入預測參數之一。



圖 3.4 台北 YouBike 捷運市政府站星期一借車數量比較圖

### 3.2.3 起訖點趨勢

在 3.2.1 小節有提及台北 YouBike 在平日時使用通勤的比例比較高，以通勤為主的話，移動的某些起訖點組合可能會有固定的發生機率，為了得知各站點的起訖點移動情況，我們計算每個站點每日的起訖點總數，如圖 3.5 顯示的是 2014 年 1 月台北 YouBike 公館站的起訖點趨勢，圖中僅列出數量較多的前幾名訖站，可以發現從公館站出發到台灣科技大學站是之中最頻繁的起訖站。有了所有站點的起訖點分配後，則可以利用起訖點的數量計算從某站移動至相對應訖站的機率值，這個機率值則可以用以預測民眾較有可能還車的站點，這個機率值會在第四章的數學模式中用到。

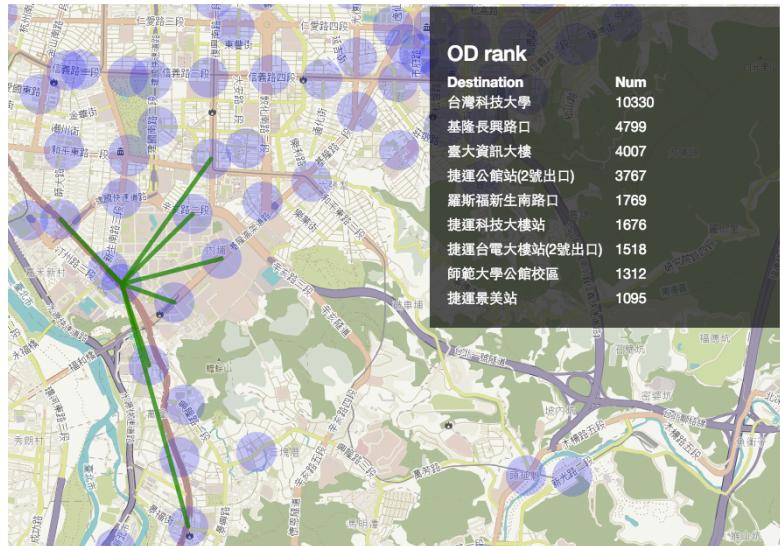


圖 3.5 台北 YouBike 2014 年 1 月公館站起訖點統計圖

### 3.2.4 時間區間

在資料分析時，會面臨到的問題是資料時間點切割的方式，圖 3.6 採用的是每五分鐘切割結算一次該時段內的總次數。若將時間區間大小拉大如圖 3.7 以每 30 分鐘結算一次，仍可以看出使用趨勢；但若以每 360 分鐘計算一次如圖 3.8 所示，則整體趨勢變化會變得不明顯，無法看出早上有一個交易高峰；反之，若時間點切的太細，會造成交易數波動甚大。有些文獻 (Montoliu, 2012) 會彈性地調整一天當中的某些時間區段長度，不見得所有站的所有時段長度皆相同。本研究會在第四章的數學模型中以不同的資料切割方式跑不同數據，驗證哪種時間區間比較適當。

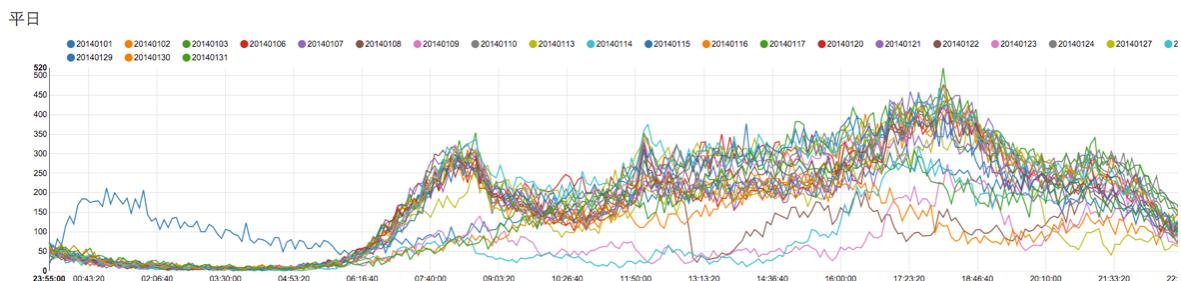


圖 3.6 每 5 分鐘結算一次

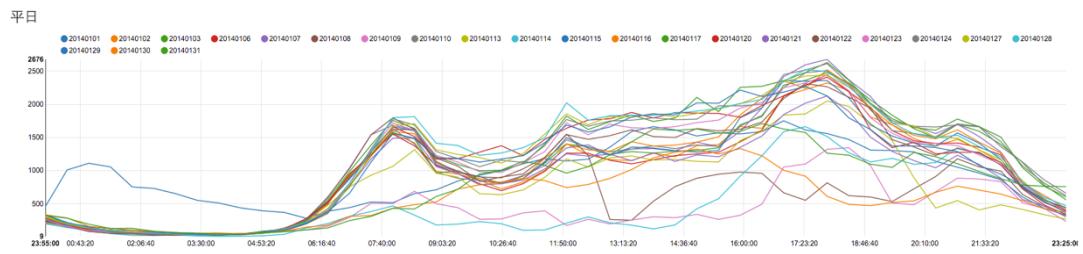


圖 3.7 每 30 分鐘結算一次

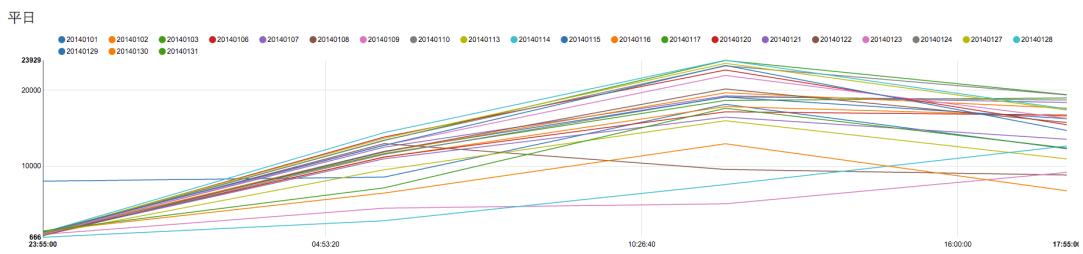


圖 3.8 每 360 分鐘結算一次

### 3.3 視覺化網頁工具

本研究將分析資料的過程包成一個程式，希望只要同為公共自行車共享系統的資料只要經過此程式處理，即可以視覺化方式呈現租借資料，並將結果呈現在網站上，此工具擬呈現使用者在站點間移動的關係、各站點的每小時借還趨勢、以及每星期的交易趨勢等等。

#### 3.3.1 使用技術

此視覺化工具以網頁呈現，使用到大量的網頁技術，以下將簡介各項技術：

- (1) HTML / CSS : HTML(HyperText Markup Language) 是一種用於建立網頁的標準標記式語言。CSS(Cascading Style Sheets)為美化 HTML 介面呈現的語言，用於將網站排版、介面優化呈現。
- (2) React.js : 由 Facebook 公開出來的開源前端框架，React.js 的框架概念為單向資料流，由於資料之間在計算會有關聯性，為了單純化介面更新的資料流，因此選用此框架。

- (3) D3.js：是一個強大的資料視覺化函式庫，在視覺化圖表上的折線圖、站點間連線皆是用此函式庫應用。
- (4) Leaflet：是一個以手機介面優先的地圖。
- (5) Ruby：此視覺化工具為事先將所需要的資料算出來，在網站上並不會做大量的資料計算，因此會將網站所需要的資料以 Ruby 語言撰寫，經由此程式算出所需要的資料並以網站呈現。

### 3.3.2 工具操作說明

此視覺化工具的目的為可以將各種不同的自行車共享系統資料，以一致的視覺化方式呈現。其操作流程如圖 3.9 所示，只要將不同系統的交易資料及地理資料放入資料處理程式，即可將分析出來的結果串接至視覺化網站。其中交易資料的格式採取.csv 格式，需提供「借車時刻、還車時刻、借車場站、還車場站」這四個欄位；地理資料要採用 json 格式，圖 3.10 即為台北 YouBike 的地理資料範例，需要提供該站點的編號、名字、經緯度以及容量。只要輸入這些資訊，系統即可自動產出視覺化結果。

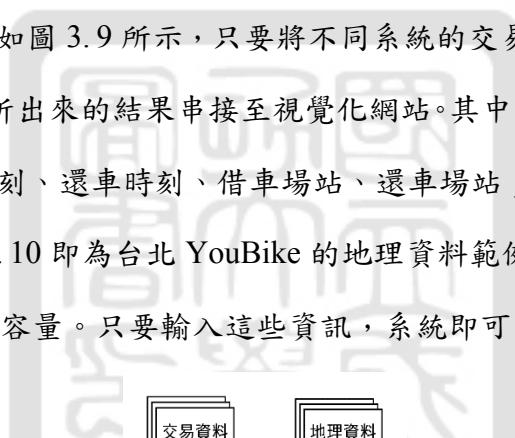


圖 3.9 視覺化工具操作流程

```
{
  "1": {
    "station": "捷運市政府站(3號出口)",
    "lat": 25.0408578889,
    "long": 121.567904444,
    "capacity": 180
  },
  "2": {
    "station": ["捷運國父紀念館站(2號出口)", "捷運國父紀念館站"],
    "lat": 25.041,
    "long": 121.556945,
    "capacity": 48
  }
}
```

圖 3.10 地理範例資料

### 3.3.3 資訊呈現

此視覺化工具將先呈現大範圍的整個 PBSS 層面的數據，再呈現小範圍的各站點個別租借數據，以下一一說明所會呈現的內容。

#### 3.3.3.1 公共自行車共享系統資訊

系統層面的資訊想要呈現的是大方向的內容，如圖 3.11 所示，所呈現的有一段期間內該系統的交易量波動、不同星期的交易總數、不同時間區段的交易總數以及該系統借車、還車最頻繁的幾個站點。經由系統層面的資料，可以初步對這個系統有大方向的了解，大致理解這個系統的交易狀況。



圖 3.11 台北 YouBike 系統視覺化範例截圖

### 3.3.3.2 站點分析—熱點圖

對於站點數較多的 PBSS 而言，難以透過單一的站點分析來了解究竟是哪些範圍是比較熱門的交易區域，因此透過熱點圖呈現，可以很直覺地看出該系統的熱門、冷門交易區域。如圖 3.12 為台北 YouBike 的熱點圖，可以從圖中看出在 2014/1/3 17:00 時的熱門交易站點位於哪些區塊。

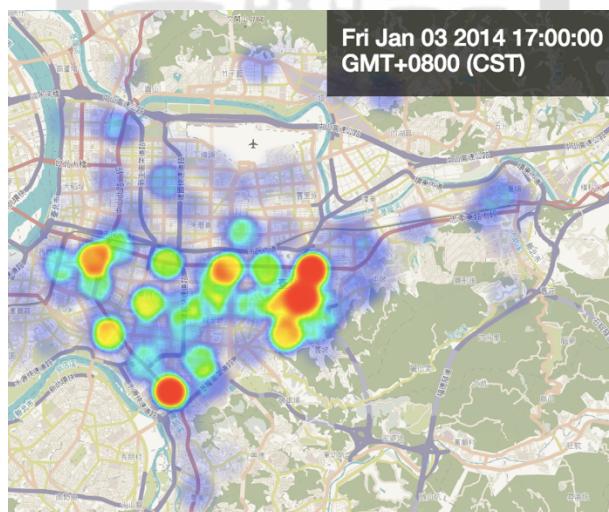


圖 3.12 台北 YouBike 站點熱點圖截圖

### 3.3.3.3 站點分析—站點資訊圖

上一小節的熱點圖呈現的是「區域型」的交易，此小節說明的是以各站點詳細交易資訊、站點間交互關係為主的視覺化呈現。如圖 3.13 為台北 YouBike 站點資訊的

截圖，主要以地圖為呈現，使用者可以點選地圖上的站點圖標，切換想要看的資訊。選取某個站點後，左側欄為該站點的整體借車、還車交易資訊，右側欄為該站點與其它站的起訖點關係列表。以下將一一解釋各個圖表所呈現的詳細資訊。

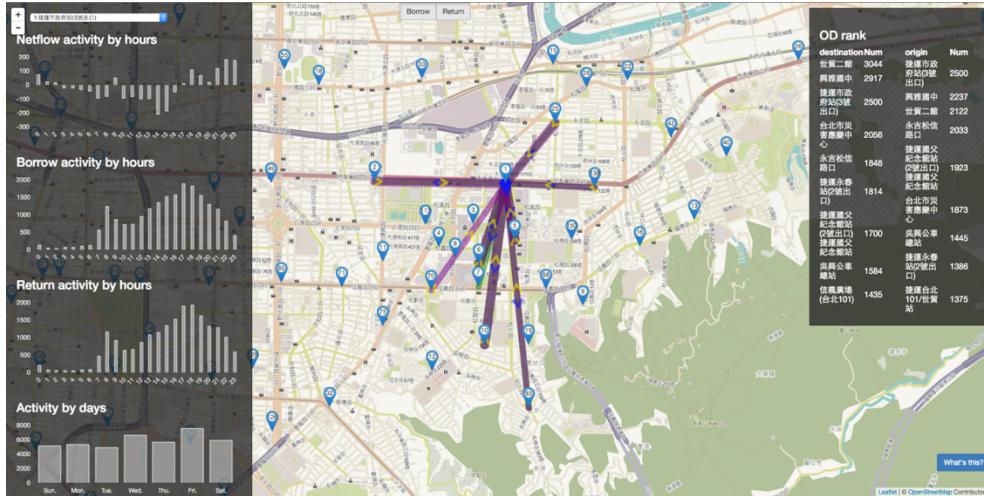


圖 3.13 台北 YouBike 站點資訊截圖

### (1) 地圖圖標

如圖 3.14 所示，此工具以地圖為基本呈現，使用者點選不同的站點，切換想要看的站點資訊，圖標中的站點編號沿用台北 YouBike 的編號方式，點選站點後則可以呈現相關資訊。

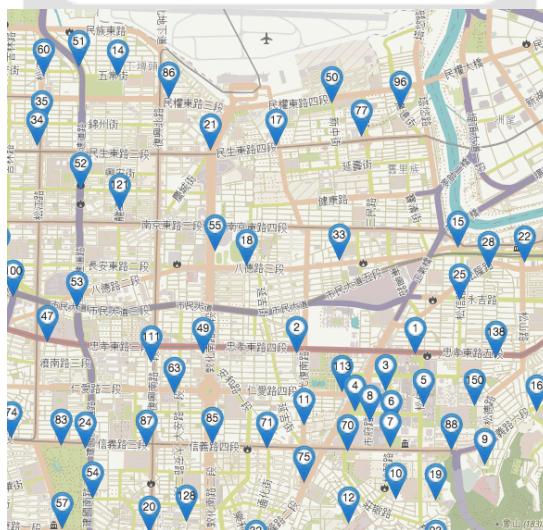


圖 3.14 視覺化工具地圖圖標截圖

## (2) 站點資訊

如圖 3.15 所示，其為捷運市政府站的站點資訊圖，第一張圖呈現每小時的淨流量（還車－借車）；第二張圖為每小時的借車數量；第三張圖為每小時的還車數量；第四張圖為不同星期的交易活動數量，一次的借與還車記為一次交易活動。經由此呈現即可了解該站點的借車還車波動以及是在哪個星期有比較多的交易量等記錄，即可透過此圖對該租借站的交易量變化有初步的認識。

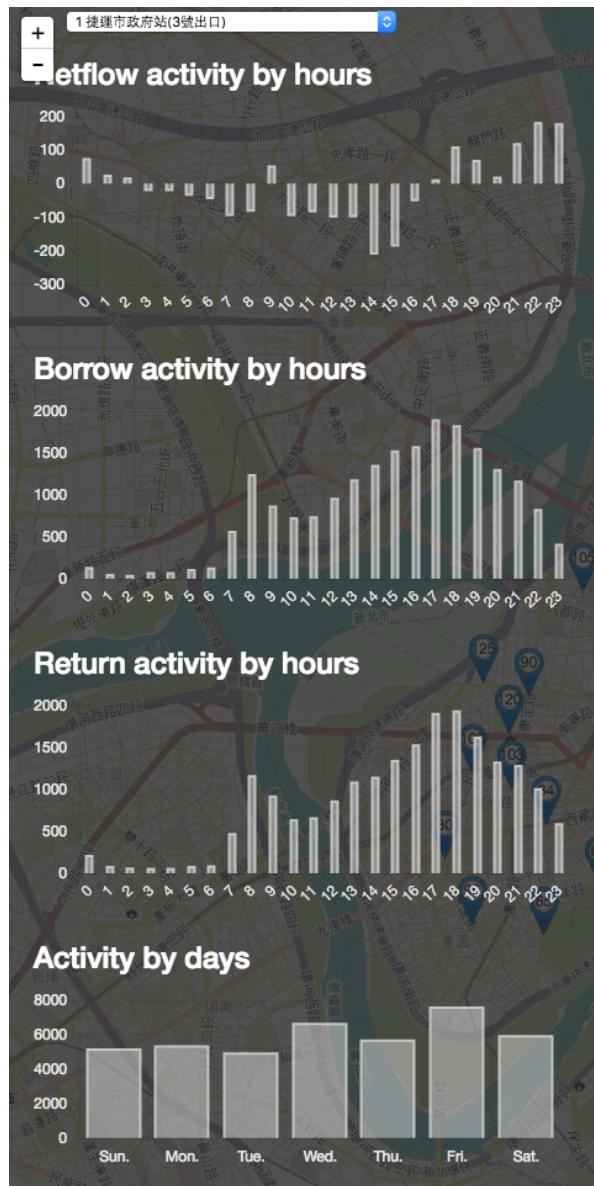


圖 3.15 台北 YouBike 捷運市政府站站點資訊圖

### (3) 站點間關係

站點間關係要呈現的是各站點間的關聯性，例如從台北火車站借車的人會比較傾向往哪些站點移動，如圖 3.16 所呈現的是捷運市政府站至各站點的起訖點關係，線的粗細代表的是該起訖流量的大小，線越粗表示越多人移動到該站點，反之則越少。圖 3.17 為捷運市政府站至其它站點的前 10 名起訖站總數，左邊為從該站點至其它站點的移動總數，右邊為其它站點移動至該站點的總數。藉由此二圖即可了解各站點至其它站點的起訖關係。

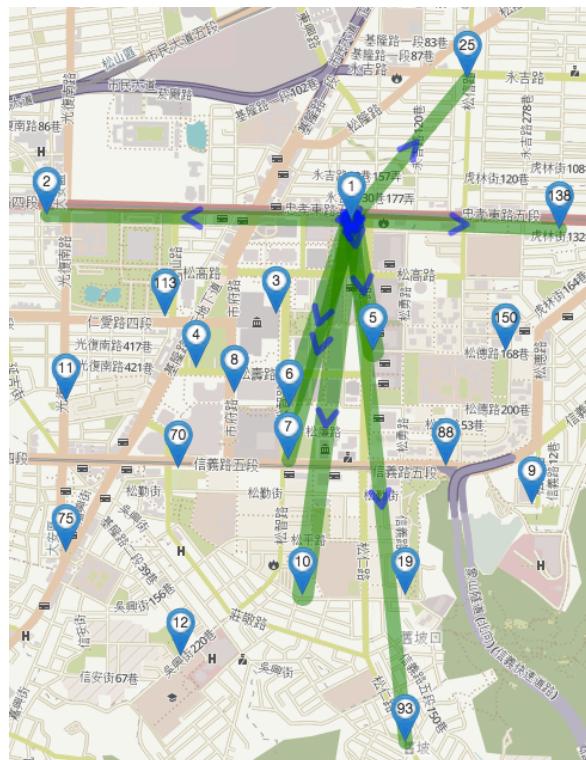


圖 3.16 台北 YouBike 捷運市政府站起訖關係資訊圖



圖 3.17 台北 YouBike 捷運市政府站的起訖旅次總數圖

### 3.3.4 視覺化網頁工具小結

本視覺化工具旨在讓想了解 PBSS 營運狀況的人能很快看出該系統的基本特性，尤其有助於系統管理員更全面、深度地了解各站點的庫存變化以及使用者的租借趨勢，對之後的運補決策更有直覺。而除了台北 YouBike 的視覺呈現，本研究也將此工具套用至諸如高雄 CityBike、紐約 CitiBike 等其它 PBSS，可前往 <http://youbike.ctxhou.com/> 檢視此工具的呈現結果。

## 3.4 小結

本章介紹 PBSS 的資料、分析方式、以及視覺化工具，以台北 YouBike 的租借資料為主要範例。資料分析結果顯示當日的型別（平日、假日、國定假日）以及天氣狀態（晴、微雨、大雨）會直接影響民眾騎乘自行車，切割時段的區間大小亦會影響數

據的呈現結果。本研究將每筆租借的相關資料依表 3.3 儲存。有了基本的分類後，當要處理數據時則可以將資料依據不同類別歸類，例如每 30 分鐘平日晴天、每 20 分鐘假日微雨等。從起訖旅程趨勢分析可得知某些站與站間的旅程量有規則可循，此起訖站移動的機率將用在第四章的群眾運補數學模式中運算。

藉由更了解資料特性，能夠幫助我們在運算時選擇相同類型的資料。最後本研究也以分析資料後的心得，將此流程包裝成一個視覺化網頁工具，以方便未來對其它公共自行車共享系統的營運狀況能在短時間內看出大致輪廓。

表 3.3 資料分類

日			天氣			時刻		
平日	假日	國定假日	晴天	微雨	大雨	時段	小時	分鐘

## 第四章 群眾運補之研究

本章會介紹群眾運補的研究，以圖 4.1 表示方法流程，本研究提出的「當期群眾運補數學模式」以能夠處理即時資料為目標，對於需要及時決定運補方式的狀況，此模型所需要的參數必須事先計算好，先用「最適車數數學模式」和「需求預測模型」來估算所需要的參數，再將這些參數放入當期群眾運補數學模式運算。最適車數數學模式會算出每個時段每站的最適車數，需求預測模型即會預測下一期各站點的需求量，計算完的最適車數及需求量再提供「當期群眾運補數學模式」運算。4.1 節介紹最適車數的數學模式；4.2 節介紹預測未來借車需求的方法；4.3 節為當期群眾運補的數學模式。

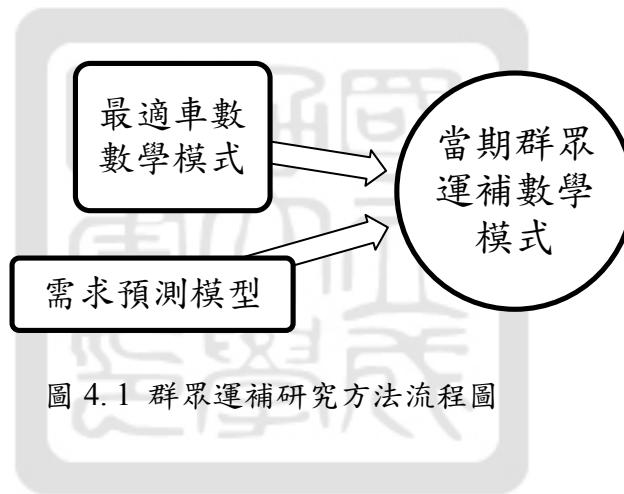


圖 4.1 群眾運補研究方法流程圖

### 4.1 最適車數數學模式

#### 4.1.1 問題描述

PBSS 在營運時，為了讓運補車在運補時有一個運補目標值，現行的作法為各站在各期末時刻訂定一個車數目標值，希望若該站能在該期末前達到該車數，則當期或之後的缺車或缺位情況會改善最多。最適車數的計算需要考量以下項目：各站點容量、全日交易資料、各站點座標位址、各站點容量用於給予志願運補的上限值等資料，其中各站點座標位址可限制志願運補所能涵蓋的範圍（譬如我們可限制志願運補的起訖站距離不能超過 2km 或騎乘時間不能超過 25 分鐘等）。

首先先定義借車與還車的時刻，如圖 4.2 所示， $t=1$  表示的是從時刻 0 到 0.9999 之間，即將近時刻 1 的這段期間，因此在  $t=1$  期末還的車，在  $t=2$  期初即可被租借。計算最適車數的概念以表 4.1 的例子來說，租借站 A 在第一期有借車 4 台、還車 2 台，亦即在第一期時，租借站 A 至少要有 4 台車才能滿足第一期的需求，而到第二期時租借站 A 有借車 3 台，但因第一期時有 2 台自行車被歸還，因此在第二期租借站 A 只需要至少 1 台車則可以滿足此期需求，以此方式類推則可以去推算在每期初（或末）、每站至少需要幾台車才可以滿足該期的最低需求，可用圖 4.3 表示站點 A 的情況，藉由此計算方式，則可以得知在各期初各站的最低車數需求。



圖 4.2 時段定義

表 4.1 範例租借站借還資料

	$t=1$		$t=2$		$t=3$	
借車/還車	借車	還車	借車	還車	借車	還車
A	4	2	3	1	1	2
B	2	3	1	2	3	1
C	1	4	2	1	2	2

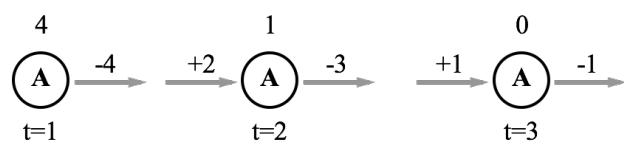


圖 4.3 範例租借站 A 之借還資料示意圖

此模型與過去研究不同之處在於，過去研究多以「站點距離」作為運補範圍的限制，使用站點距離的缺點在於目前站點距離多只能使用直線距離做估算，但實際道路為彎曲路線，因此直線距離一定比實際距離來的短，估算上較不準確。此外，以距離做判斷是建立在群眾騎乘自行車是以站點間移動為目的，但如高雄 CityBike 目前實施前一小時免費的機制，有許多民眾會將騎自行車當作運動，騎乘時間遠比騎乘直線距離來的長。另外，如圖 4.4 表示的是台北 YouBike 的國興青年站訖點總數，可以發現此站的騎乘特性為許多交易都是同起訖點，如此無法用單純的直線距離來計算。所以本研究不以距離為運補範圍限制，而改以計算「騎乘時間」。過去的研究由於資料不足等原因，無法以使用者實際的騎乘時間來估算站與站之間的騎乘時間，然而本研究可以藉由分析 YouBike 騎乘資料估算各時段各起訖點之間的移動時間；同理，若歷史資料中從未發生某組騎乘起訖旅程的話，該旅程將被視為不可能發生。



圖 4.4 台北 YouBike 2014 年 1 月國興青年路口站訖點總數

#### 4.1.2 問題假設

針對本問題的情境，列出下列假設：

1. 對每一時段而言，借車在期初、還車在期末發生。
2. 所有租借者的交易資料都已知。
3. 可以號召無限多的群眾。

4. 使用者起訖站的騎乘時間為已知常數。

### 整數規劃模式

#### 參數

$N$  表示總車站數，租借站編號  $i, j = 1, \dots, N$

$T$  總時間期數，時間  $t = 1, \dots, T$

$U_i$  表示第  $i$  站的容量值

$b_i^t$  表示在第  $i$  站在第  $t$  期期初所借出的車數

$r_i^t$  表示在第  $i$  站在第  $t$  期期末所歸還的車數

$\varepsilon$  為一個極小的數字

$t_{ij}$  表示從第  $i$  站到第  $j$  站騎乘自行車所需花的時間期數

#### 變數

$hb_{ij}^t$  表示第  $t$  期期初從第  $i$  站前往第  $j$  站採取群眾運補所借的車數

$hr_{ij}^t$  表示第  $t$  期期末從第  $i$  站前往第  $j$  站採取群眾運補所還的車數

$\Delta U_i^t$  表示在第  $t$  期期末時，因滿站而無法於租借站  $i$  還車的數量

$\Delta L_i^t$  表示在第  $t$  期期末時，因空站而無法於租借站  $i$  借車的數量

$I_i^t$  表示第  $t$  期期末時，第  $i$  站的自行車數

上述參數中，其中總時段數  $T$  的可依訂定的時段長度（譬如 5、15、30 分鐘等）而改變，而依時段長度可定出移動時間  $t_{ij}$  的期數值，舉例來說，以表 4.2 所示，從站點 A 到 B 的實際騎乘時間是 11 分鐘，若以 5 分鐘為一個區間的話，其騎乘時間視為 3 期；若以 15 分鐘一個區間則視為 2 期，而以 20 分鐘一個區間則視為 1 期。當然，若時段長度太長，則較為不準確。

表 4.2 時間點切割範例

	實際騎乘時間	5 分鐘切割	15 分鐘切割	20 分鐘切割
站點 A → 站點 B	11 分鐘	3 期	2 期	1 期

#### 4.1.3 數學模式

最適車數數學模式要計算的是在已知全日的交易量下，如何最佳的分配各時段各站的最適車數，目標值如式 4.1.1 所示，其中  $\Delta U_i^t$  及  $\Delta L_i^t$  表示有缺車或是缺位的需求不滿足數量，亦即目標為最小化總需求不滿足的人次；權重  $\varepsilon$  為一個極小的常數， $hb_{ij}^t$  代表的是群眾運補的數量（以借車的角度），將  $\varepsilon \sum_{t=1}^T \sum_{i=1}^N \sum_{j=1}^N hb_{ij}^t$  放入目標式是為了避免無謂的群眾運補（譬如指揮兩位志願騎乘者在 A、B 兩站間連續兩期互相運補：A-B-A, B-A-B，雖可滿足限制式但其運補效果互相抵消沒效益）。

$$\min \sum_{t=0}^T \sum_{i=1}^N (\Delta U_i^t + \Delta L_i^t) + \varepsilon \sum_{t=1}^T \sum_{i=1}^N \sum_{j=1}^N hb_{ij}^t \quad (4.1.1)$$

每一期的期末總車數為前期期末總車數減去當期期初的原使用者及群眾運補之借車數，加上前期期末的原使用者及群眾運補之還車數，最後加上需求未滿足量，如式(4.1.2)。

$$I_i^t = I_i^{t-1} + \sum_{j=1}^N (-b_{ij}^t + r_i^{t-1} - hb_{ij}^t + hr_{ji}^{t-1}) - \Delta U_i^t + \Delta L_i^t \quad (4.1.2)$$

$$\forall t = 1, \dots, T; i = 1, \dots, N; j = 1, \dots, N$$

群眾運補的車數需大於等於零，如式(4.1.3)。

$$hb_{ij}^t \geq 0 \quad \forall t = 1, \dots, T; i = 1, \dots, N; j = 1, \dots, N \quad (4.1.3)$$

禁止同一起訖點的群眾運補，如式(4.1.4)。

$$hb_{ii}^t = 0 \quad \forall t = 1, \dots, T; i = 1, \dots, N \quad (4.1.4)$$

當在第  $t$  期時有指派從第  $i$  站到第  $j$  站的群眾運補，則在經過其騎乘時間  $t_{ij}$  期後，預期這些群眾運補的車會被歸還至第  $j$  站，如式(4.1.5)。

$$hr_{ij}^{t+t_{ij}} = hb_{ij}^t \quad \forall t=1,...,T; i=1,...,N; j=1,...,N \quad (4.1.5)$$

各期間的各站車數不可大於其最大容量，如式(4.1.6)。

$$I_i^t \leq U_i \quad \forall t=1,...,T; i=1,...,N \quad (4.1.6)$$

簡言之，此數學模式假設一個有無限多志願騎乘者可協助運補的理想情境，因為使用者的起訖旅程都已知，且號召志願騎乘者又不費成本，因此得到的每站每期期末的自行車個數 $I_i^t$ 應該是一個最佳的理想值，而該值即可當成當期群眾運補的運補目標。

## 4.2 預測未來需求

預測未來需求中，未來借車需求相較於還車來說不確定性更高，借車的意願可能會受到當時天氣、氣溫、車站數、交通狀況、站點週邊商家數、人群數等等因素影響，如 Conway (2014) 則將 7 個變數放入線性回歸模型。預測未來借車需求的作法主要有兩種方向，其一是假設各租借站的特性相似，因此可用同一方式來預測任一站發生的租借趨勢；其二則假設各租借站皆為獨特的個體，針對每個站點產生個別的預測模型。本研究採取後者，因為分析過台北 YouBike 的租借資料後發現各站的特性不一，譬如在觀光地、學校、大車站、公園等地的租借需求表現各有不同。以下將敘述本研究曾經嘗試過卻失敗的「線性回歸」與「時間序列分析」等兩類預測模型。

本研究將多個可能會影響台北 YouBike 的變數放進線性回歸模式，跑出其回歸模型，但測試多個站點後發現其預測結果的殘差值表現皆不理想，會有誤差過大的情況發生，此外其調整後的 $R^2$ 結果多介於 0.5-0.6 之間，表示其模型的解釋力不佳，因此並未選用線性回歸。

時間序列分析常應用在金融業、財務分析領域，本研究嘗試以 ARIMA 模型建構出台北 YouBike 的時間序列模型，並預測未來可能的借車還車值，但因台北 YouBike 的交易波動起伏會受到許多參數影響，我們難以將資料準確地依據不同的參數組合分

類並放進各自的時間序列預測，若僅是以幾種參數組合作預測，最後結果並不佳，因此不選用時間序列方法。

本研究最終選擇以「隨機森林演算法」預測未來借車、還車需求，其能處理大量資料以及考量多種參數組合的特性適用於本研究的狀況。其中，訓練資料集、特徵值、決策樹的數量以及每次最少選幾個特徵值建立一棵決策樹等皆可能影響隨機森林演算法的表現。在 5.3 節隨機森林預測模型會分別針對影響隨機森林演算法的參數做測試，選出最適合台北 YouBike 的隨機森林模型。以本研究採用的台北 YouBike 資料共有 164 個站，將針對不同站分別產生不同的隨機森林模型，依據當下所要預測的站以及相關參數，預計下一個時段的借車和還車量，並利用此預測值輔助運補的決策。

### 4.3 當期群眾運補數學模式

PBSS 的日間需求變化十分動態，僅用夜間靜態運補的方式難以平衡一整天即時的狀況，因此本研究從「動態運補」的方向出發，提出群眾運補策略及模式。

#### 4.3.1 當期群眾運補策略

本研究針對群眾運補模型採取的是單期的運補策略，即此模型在當下僅會決定當下那一期的運補策略，僅考量當期的原因在於動態運補屬於即時情況，盡量達到當前即將發生的需求最為優先。欲做群眾運補，首先要先計算當期期末各站的自行車數，因此需要考量的點有 (1)已借出但尚未歸還的自行車；(2)未來預期會被借出的自行車，其中未來預期會被借出的自行車將使用 4.2 節及 5.3 節的隨機森林模型預測之。

##### 4.3.1.1 已借出但尚未歸還的自行車預測方式

在動態運補的情況下，考量的是當下至下一期之間的情況，但如圖 4.5 所示，在這段期間可能會有在此時段之前租借的車尚未被歸還，因此需要考量這些資料。

計算方式為利用起訖點移動機率計算，在3.2節資料分析有提及站點的起訖點移動有一定的趨勢，所以可以利用歷史資料計算在各時段各站點的移動趨勢，接著再以實際資料估算預期會歸還車數。以圖4.6為例，A點在某段期間內的起訖點移動資料如圖所標示：2筆A至B、5筆A至C、1筆A至D、5筆A至E。經由從歷史資料分析出交易起訖點的資料則可以歸納出表4.3，A站至各點的機率，將此計算方式延伸至所有站點，因此每站每期、不同天氣都會有其起訖點的移動機率。

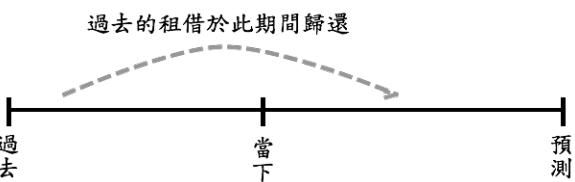


圖4.5 已借出但尚未歸還之車流量示意圖

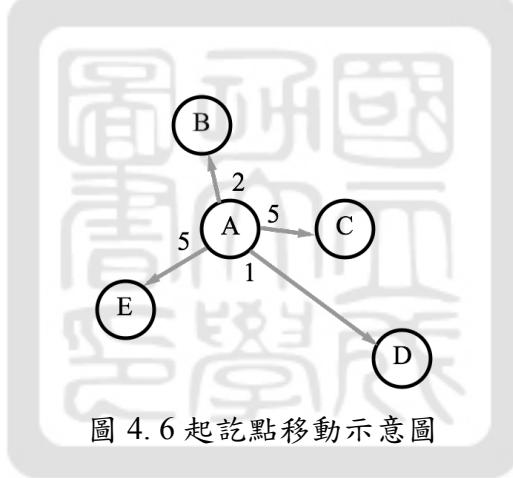


圖4.6 起訖點移動示意圖

表4.3 範例A站起訖點移動機率

	B	C	D	E
A	0.16	0.38	0.08	0.38

表4.4 範例A站租借10台車的訖點車數估算

	B	C	D	E
A	1.6	3.8	0.8	3.8

從歷史資料計算出各站點的移動機率後，則可以去接台北 YouBike 的即時站點資訊以及氣象局的即時天氣資訊，藉由當下的時刻和天氣狀況，選擇當時的起訖點機率計算。台北 YouBike 的即時資料為每 15 分鐘更新一次，我們可以依據站點在這 15 分鐘間的差值去計算有幾台車被租還，再依據這個差值去算被借走的車子會移動到各站的機率為何。舉上述 A 站的例子來說，若 A 站在這 15 分鐘間有 10 台車子被借走，則可以依據機率做等比例分配，計算移動至各站點的車數，如表 4.4。所示，此數字則會放入數學模型做群眾運補的推算。

#### 4.3.2 整數規劃模式

當期群眾運補整數規劃模式的輸入資料為當期已知的參數如下：

##### 參數

$A$  表示該期所能號召幫忙騎乘的志願者人數上限

$N$  表示總車站數，租借站編號  $i, j = 1, \dots, N$

$L_i$  表示第  $i$  站期末的最佳車輛數目標值，即若達到此值，則可以比較容易達成未來的需求

$U_i$  表示第  $i$  站所能放的最大自行車數

$I_i$  表示第  $i$  站的期初自行車數

$b_i$  表示該期內從第  $i$  站借出的自行車數

$r_i$  表示該期內歸還至第  $i$  站的自行車數

##### 變數

$x_{ij}$  表示此期從第  $i$  站騎往第  $j$  站採取群眾運補的車數

$k_i^+; k_i^-$  計算目標式絕對值時的代數

## 數學模式

目標為當結束該期群眾運補後，期末時的自行車數與先前各站各期的最佳車輛數（4.1.3 節）之差距最小，如(4.3.1)所示，但因為差距必為正值，所以將目標式放入限制式(4.3.7)，並將目標式轉化成(4.3.2)。

$$\min \sum_{i=1}^N \left| L_i - \left( I_i + \sum_{j=1}^N (-x_{ij} + x_{ji} - b_i + r_i) \right) \right| \quad (4.3.1)$$

$$\min \sum_{i=1}^N (k_i^+ + k_i^-) \quad (4.3.2)$$

因為不可能每次系統號召的人數皆可剛好滿足，所以訂定  $A$  為可號召到的人數最大上限，即最多只能找到  $A$  人執行志願運補，如式(4.3.3)

$$\sum_{j=1}^N \sum_{i=1}^N x_{ij} \leq A \quad \forall i = 1, \dots, N; j = 1, \dots, N \quad (4.3.3)$$

志願運補之後的各站車輛數仍須符合各站最大容量，如式(4.3.4)

$$0 \leq I_i + \sum_{j=1}^N (x_{ji} - x_{ij}) + r_i - b_i \leq U_i \quad \forall i = 1, \dots, N; j = 1, \dots, N \quad (4.3.4)$$

因為此模型考量的是短時間的當期運補，因此忽略同站短期借還的可能性，即志願運補的流量不能起訖點相同，如式(4.3.5)

$$x_{ii} = 0 \quad \forall i = 1, \dots, N \quad (4.3.5)$$

群眾運補的流量必為正或 0，如式(4.3.6)

$$x_{ij} \geq 0 \quad \forall i = 1, \dots, N; j = 1, \dots, N \quad (4.3.6)$$

此模型的目標式為最小化志願運補後之車數與該期期末之最佳車輛數差距，但因其包含絕對值，因此將目標式移到限制式定義如式(4.3.7)

$$k_i^+ - k_i^- + I_i + \sum_{j=1}^N (-x_{ij} + x_{ji}) - b_i + r_i = L_i \quad \forall i = 1, \dots, N; j = 1, \dots, N \quad (4.3.7)$$

$k$  值必為正或 0，如式(4.3.8)

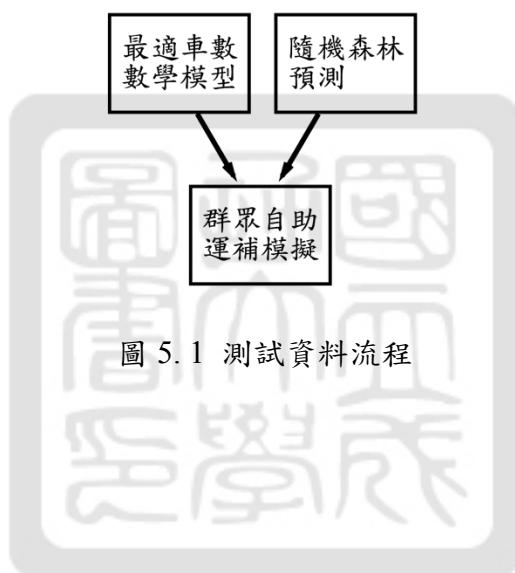
$$k_i^+, k_i^- \geq 0 \quad \forall i = 1, \dots, N \quad (4.3.8)$$

#### 4.4 小結

本章介紹當期群眾運補的相關數學模型，當期群眾運補模型以能處理即時資料為目的，利用最適數學模式和借車預測模型事先求出當期群眾運補模型所需要的參數。最適數學模式算出來各期間各站的目標車數，意義為若各站在該期期末時皆能達到此目標車數，則預期能降低系統缺車或缺位的問題。過去的研究中，動態運補僅從已知的靜態資料來計算整日的最佳運補方式，無法因應現實資料動態變化來重新估算，現行作法只能倚賴過去個人經驗或是歷史資料去猜測未來各站可能的借車人數，本研究用借車預測模型求得下一期預期的借車數輛，隨著可參考的歷史資料越多，此預測模型亦應會越準確。利用事先算好的目標車數以及預測模型，可以簡化實際在計算運補時的運算時間，此作法與過去文獻相比在面對即時動態運補時，不僅更符合現實情況，也可以大幅減少在運算時的變數，增加即時動態運補的效率與效能。

## 第五章 實際範例與數值分析

本章將會針對第四章所提及的數學模型及隨機森林預測模型做數值分析，其中數學模型的資料會用到第三章資料分析的結果。5.1 節會說明測試環境，以及真實資料如何應用於數值分析；數值分析的測試流程如圖 5.1 所示，首先 5.2 節為算出各期的最適車數數學模式，此數值即為群眾運補模擬的各期期末目標車數；5.3 節為隨機森林預測模型，用以預測當下那一期即將出現的租借需求總數；5.4 節為群眾運補模擬，會比較群眾運補與傳統卡車運補的效益；最後於 5.5 節總結數值分析結果。



### 5.1 測試情境

本研究以台北 YouBike 2014 年 1-10 月的歷史資料為測試資料集，平均每個月有約 140 萬筆交易量，每筆交易格式如表 3.1，每筆資料即為一次的借還車事件。

如第三章所述，我們可以將資料依據日、天氣以及時刻來區分，在數值分析部份為了避免資料本質不同，導致分析產生偏誤，本研究會將同樣屬性的資料放在一起，以同樣屬性的資料訓練隨機森林模型以及模擬測試，舉例來說平日晴天、平日微雨、假日晴天、假日微雨等，針對不同的情境會套用相對應的資料測試。

本研究測試的電腦處理器為 Intel(R) Core(TM) i7-4770 CPU @ 3.40GHz，作業系統為 Ubuntu 14.04.4 LTS，使用 Ruby 程式語言處理 YouBike 資料以及模擬程式；R 用

作隨機森林模型建制；C++與 Gurobi 結合求解整數規劃模式；Javascript 等網頁語言用以資料視覺化。

## 5.2 最適車數數學模式

本小節將以台北 YouBike 的歷史資料，利用最適車數模型計算出一日中各期期末的最適車數值。由於每日的交易狀況不同，本研究採用多日的最適車數平均，並設定該值為模擬測試的各期目標車數。

表 5.1 為利用 Gurobi 針對每日（譬如 3/10, 11, 12,...,etc.）租借資料求解最適車數模型得到的範例結果，每日的每期（譬如 5:00, 6:00,..., etc.）皆可由模式算出該期期末的最適車數值，在 3.2 節曾提及會將資料依照平日、假日、國定假日來區分，因此會依據平日和假日分別去計算其最適車數。表 5.1 為平日最適車數的範例值，表中為台北 YouBike 最大的車站—捷運市政府站，依據每天每期得出的最適車數，平均求出此站在各期期末的目標車數值。用此方法套用至各租借站，因此台北 YouBike 每一個租借站的每期都會有一個期末目標車數值，此值將用在 5.4 節的模擬測試中。

表 5.1 平日最適車數計算範例—台北 YouBike 捷運市政府站

	3/10	3/11	3/12	3/13	3/14	平均目標車數
5:00	174	180	177	150	162	159
6:00	148	151	152	145	132	138
7:00	98	102	115	112	76	91
8:00	22	53	49	148	28	41
9:00	31	55	75	145	52	57

## 5.3 隨機森林預測模型

本研究採用隨機森林演算法來預測未來借車、還車量，其中會影響隨機森林演算法的因素有訓練資料集、特徵值、決策樹的數量以及每次最少選幾個特徵值建立一棵決策樹等，本節將說明如何選出最適合台北 YouBike 的決策樹數量、特徵值組合，最後說明如何產生各站點的隨機森林預測模型。

### 5.3.1 決策樹數量

隨機森林是以多個決策樹來決定分類結果，因此決策樹的數量會影響其分類準確度，決策樹越多並不一定代表會有比較高的準確度，越多的決策樹必須花費較多的時間建制模型，圖 5.2 為用 YouBike 資料跑隨機森林的演算法，橫軸代表的是決策樹數量，縱軸為誤差值，可以看出大約在決策樹 150 棵左右時，誤差值則趨於穩定，僅有小數點的差距。

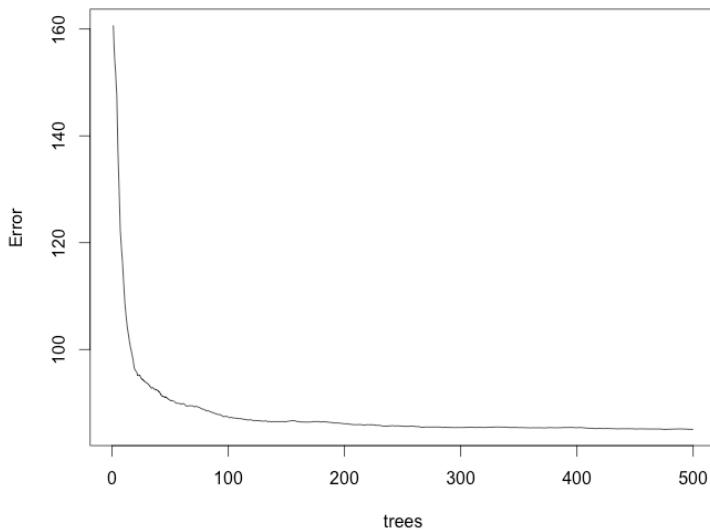


圖 5.2 隨機森林決策樹數量與誤差比較圖

針對超過 150 棵樹的預測準確率，本研究分別在同樣的參數條件下，以 500、600、1000、2000、3000 棵樹分別計算其平均方根誤差(RMSE)，結果如表 5.2 所示，當森林長到 3000 棵樹並不會使得誤差值降低許多，有時候反而會比產生 2000 棵來的差；此外產生越多樹所要花的計算時間則越長，因此在誤差值沒有差很多的情況下，我們決定選擇產生 500 棵決策樹。

表 5.2 決策樹數量的 RMSE 值比較

決策樹 數量 站點編號	500	600	800	1000	2000	3000
1	8.2605	8.2578	8.2622	8.2628	8.2624	8.2629
2	6.7989	6.799	6.7945	6.8101	6.8073	6.7997
5	4.2614	4.2604	4.2596	4.266	4.2629	4.2645
7	4.118	4.1201	4.1219	4.1262	4.1262	4.1208
10	4.439	4.4315	4.4323	4.4337	4.4415	4.4363

### 5.3.2 特徵值

針對台北 YouBike 的資料集，本研究選出九個可能會影響民眾使用需求的特徵值，分別如表 5.3 所示，其中降水量依據不同的雨量大小將其分為三個級別，如表 5.4 所示，級別越大表示其降水量越多。

表 5.3 隨機森林特徵值

特徵值	假日	工作日	降水量	溫度	前期交易數
數值	T/F	T/F	3 個級別	實際數字	實際數字
特徵值	溼度	風速	星期	小時	
數值	實際數字	實際數字	1-7	0-23	

表 5.4 降水量分級

降水量(mm)	0 – 2.5	2.6 – 8	> 8
級別	1	2	3

最適合台北 YouBike 的特徵值組合之選取方式如下：首先將所有特徵值放進隨機森林觀察每一個特徵值的重要性，結果如圖 5.3 所示，得知「當下時刻(hr)」、「前期借/還車數(lastCount)」以及「降水量(weather)」為最重要的三個特徵值，因此將以此特徵值重要度的排序從中篩選出四組特徵值組合，如表 5.5 所示，特徵值 A 為全

部皆選，特徵值 B 為篩掉風速及溼度，特徵值 C 為只選入最重要的前三個特徵值，特徵值 D 為加上假日與工作日的差別。最後我們會測試此四種組合，再從中選出 RMSE 值最低者作為台北 YouBike 的隨機森林模型的特徵值組合。

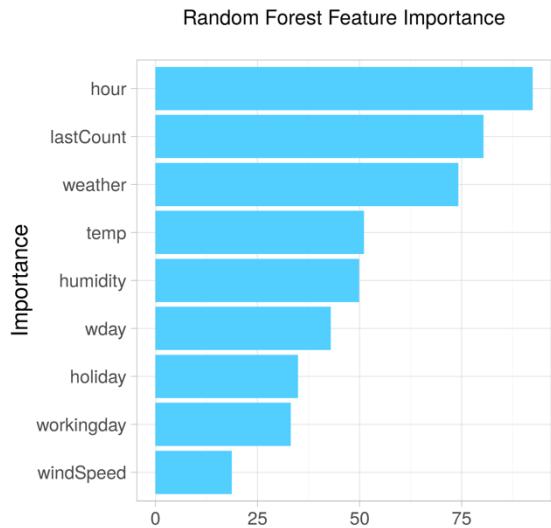


圖 5.3 特徵值重要性比較圖

表 5.5 四組特徵值組合

特徵值 A	假日、工作日、降水量、溫度、前期交易數、星期、小時、溼度、風速
特徵值 B	假日、工作日、降水量、溫度、前期交易數、星期、小時
特徵值 C	降水量、前期交易數、小時
特徵值 D	降水量、前期交易數、小時、假日、工作日

特徵值組合的測試會以同樣的決策樹數目、同樣的最低特徵值數量，搭配不同的特徵值測試五個不同的台北 YouBike 租借站。結果如表 5.6 所示，可以看出在這五個租借站中，特徵值 B 的 RMSE 值皆最小，表示其預測的誤差值最低，因此隨機森林模型將以特徵值 B 的組合來做模型的訓練。

表 5.6 四種特徵值與各站點 RMSE 測試結果

	特徵值 A	特徵值 B	特徵值 C	特徵值 D
租借站 1	10.3846	10.2867	10.9763	10.4594
租借站 2	7.0691	6.9203	7.5119	7.069
租借站 5	5.7009	5.607	5.8896	5.6379
租借站 7	5.4753	5.4105	5.6754	5.4595
租借站 10	4.9955	4.9295	5.0754	4.9991

### 5.3.3 模型建制

建制隨機森林模型時，本研究用 2014 年 1 月至 7 月的 YouBike 交易資料為訓練資料，2014 年 8 月至 10 月的交易資料為測試資料，每一個租借站以 5.3.2 節得出的最佳特徵值組合 B、500 棵決策樹，並將資料以及借車和還車分開的情形下，分別產生相對應的隨機森林模型。這些預先算出的隨機森林模型會事先儲存起來，在模擬過程中再依據當下的租借資料當成參數輸入隨機森林模型，即可得其預測值。

### 5.3.4 隨機森林小結

對於台北 YouBike 的隨機森林模型，本研究發現決策樹的數量 500 與 3000 棵的誤差值差距不大，但所花費的運算時間卻差很多（產生 3000 棵決策樹的隨機森林運算時間平均花費 30 分鐘左右），因此選擇產生 500 棵決策樹。在特徵值選擇上，特徵值 B 組合的結果最好，因此採用特徵值 B。本研究會針對台北 YouBike 164 個租借站每站分別產生一個隨機森林的模型，在模擬階段中，會依據當下對應特徵值 B 組合的相關參數，分別對 164 個已經事先算好的模型一一產生其預測值。

## 5.4 模擬測試

本研究將以模擬測試的方式驗證比較群眾運補與卡車運補的效益，以相同的租借資料但是不同的運補方式測試，模擬的流程如圖 5.4 所示，模擬一整天的租借事件後，分別再用不同的運補方式處理，最後比較兩者的服務績效。

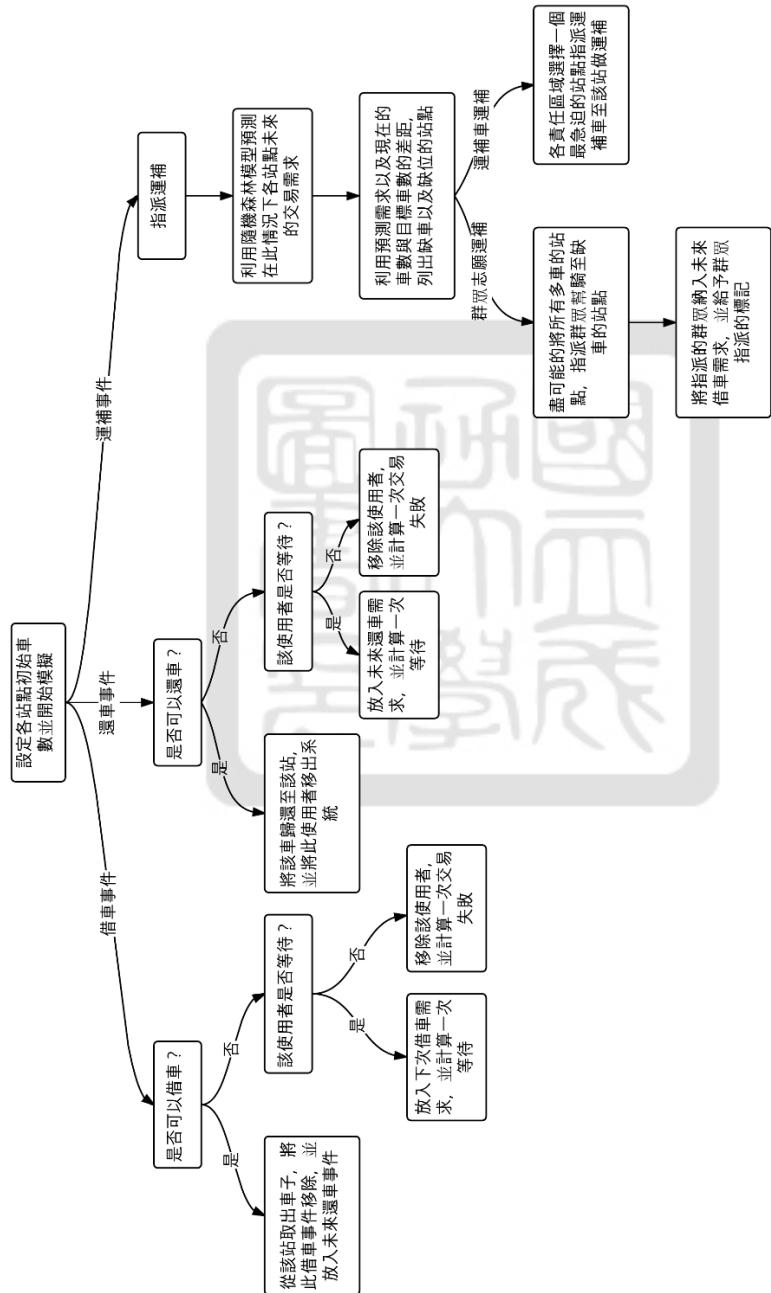


圖 5.4 模擬流程圖

模擬過程中將產生三種事件：借車事件、還車事件以及運補事件

### 1. 借車事件

當借車事件發生時，將先檢查該站點是否有足夠的車可以借，若有，則將該車移出，並將此使用者依其預期的還車時刻安插進未來事件中；若沒有足夠的車，則依據系統設定的使用者等待機率模擬是否等待，若使用者決定等待，則模擬一等待的時間，再依其預期的最晚等待借車時刻安插進未來事件中，若至該時刻仍無車可借，即認定此使用者放棄租車，累計一次租車失敗記錄；若使用者一開始即決定放棄等待，則將此使用者移出系統，並累計一次租車失敗記錄。

### 2. 還車事件

還車事件是由借車事件所引發，一遇到還車事件時，首先判斷是否可以還車，若可以還車則將該車還至該站，並將此使用者移出系統（成功還車時若該站有人正等待借車，則立即引發借車事件）；若因滿站而無法還車，則模擬使用者是否等待，若使用者決定等待，則模擬一等待的時間，再依其預期的最晚等待還車時刻安插進未來事件中，若至該時刻仍無位可還，即認定此使用者放棄還車，累計一次還車失敗記錄；若使用者一開始即決定放棄等待，則將此使用者移出系統，並累計一次還車失敗記錄。在此，我們先假設滿站時可將車子暫放在臨停區，但仍算是未成功還車，因此這些車之後仍可被租借。

### 3. 運補事件

運補事件以每隔一定時間間隔（譬如每 30 或 15 分鐘）之後所引發，此時間間隔必須在模擬前決定好。運補一開始即會依據當下偵測而得的特徵值，以隨機森林模型預測在本期間會有多少借車量，求出此值後則可以依據當下車數以及本期預計之還車數（亦由隨機森林模型預測而得），算出本期期末時刻之預期總車數與目標車數的差值，接著依據各站的缺車（即預期總車數低於目標車數）或缺位（即預期總車數高於目標車數）的狀況來決定如何運補：

### (1) 群眾運補

群眾運補由於考量到民眾在短時間內所能騎乘的範圍有限，因此會設定一個合理的運補騎乘距離範圍（譬如 2 公里），當成每趟志願騎乘運補的旅程距離上限。運補方法範例如圖 5.5 所示，首先將缺車以及缺位的租借站分為兩邊，由缺位（即較滿車）的站號召群眾幫忙騎乘至合理的運補騎乘距離範圍內的某個缺車的站，以盡量將其缺車或缺位量弭平。若可號召配對之志願騎士越多，應越能減少未來的缺車或缺位機會，因此缺車或缺位較嚴重的租借站會優先被運補，亦即優先由車較多的站去補車較少的站，直至無法再運補（譬如已達總志願運補者人數上限，或所有缺位的站皆已在在其能力範圍內運補完其可負責的缺車站）為止。

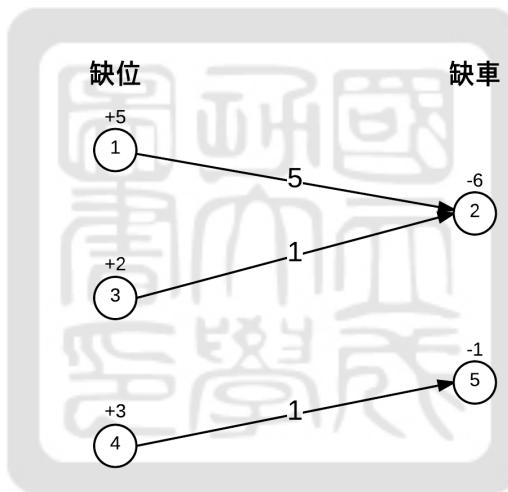


圖 5.5 群眾運補運補示意圖

### (2) 卡車運補

為更能符合現實的卡車運補作法，我們在模擬卡車運補時會先限制每台卡車的服務範圍（亦即可卸載車輛的租借站集合）。以站間距離與各站需求數利用 K-means 演算法來劃分各卡車之責任區域分群，每台運補車僅負責自己區域內站點的運補。舉例來說，圖 5.6 即為僅考慮站間距離的關係下，將台北 YouBike 租借站分為 10 羣的結果，每群皆有其專用的一台運補車協助運補，運補車的運補範圍即為該群內的所有租借站。

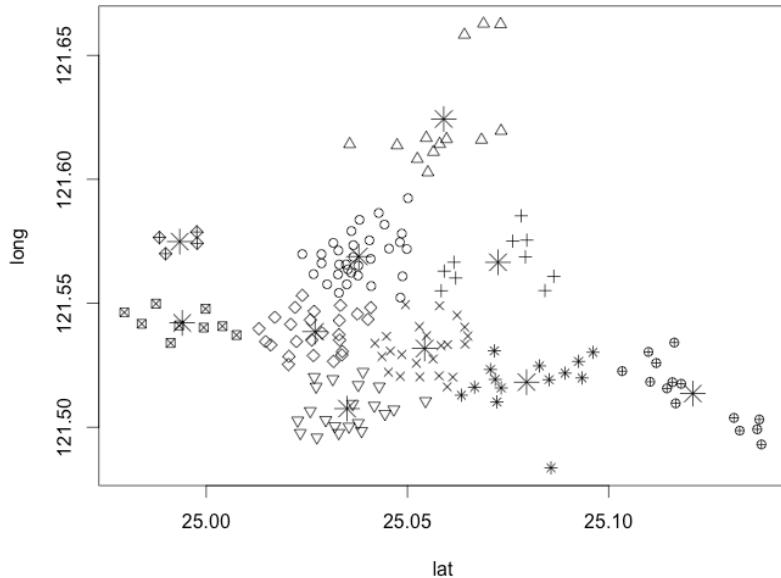


圖 5.6 台北 YouBike 以 K-means 分成 10 群分群圖

而在運補車的運補策略上，由於卡車在市區移動速度較為緩慢，平均而言移動到鄰近站點約花費 10-15 分鐘的時間，再加上運補的時間，半個小時最多僅能運補一個站點，因此運補車的策略設定成一個較簡易的貪婪法則：考量運補車當下已載的車數，在此期內運補該車所能應付的最急需的站。舉例來說，如果每台運補車的容量為 20 台自行車，如圖 5.7 左圖，租借站 5 跟 2 為目前最需被運補的站（皆需處理 10 台，然而前者為移車、後者為補車），由於當下卡車上已經裝了 18 台車，僅可再裝 2 台，因此選擇去租借站 2 運補（自車卸下 10 台車至站）；而圖 5.7 右圖則反而變為選擇租借站做運補（自站移 10 台車至車）。

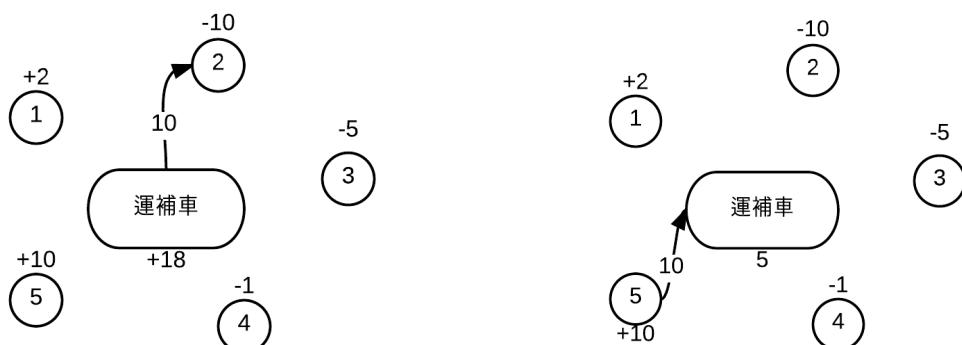


圖 5.7 運補車運補策略示意圖

模擬測試會利用 5.2 節所得出的最適車數作為各站各期期末的目標車數，用 YouBike 一日的真實交易資料輸入給模擬系統，亦即先將這些租借資料當成必定發生的借車事件，若可成功借車則隨即以該資料的還車記錄當成即將發生的還車事件，至於是否能成功借還車則視當下該站的車輛數而定；而運補事件將在每隔一段時間（譬如 30 分鐘）引發一次，依當下情況決定當期的運補決策：以群眾運補而言，就是各缺位站應找多少位志願者騎去哪個缺車站；以卡車運補而言，就是依卡車當下位置及其車上庫存量，並考慮該期內可抵達的其它租借站的缺車或缺位量，決定當下應移動至哪一站卸載多少自行車。

值得一提的是，我們模擬測試使用真實的台北 YouBike 歷史交易資料，而該歷史資料每一筆皆為成功交易的結果，未記錄那些本來應發生但卻因缺車或缺位而未成功的租借需求；同理，這些真實資料亦無法區分某筆成功的交易是否為經歷失敗後的結果（譬如該租借者可能已經歷缺車或缺位，而在等待或換租借站後才成功借還）。

模擬測試有以下假設及參數設定：

- (1) 假設沒有群眾號召人數的限制
- (2) 群眾在站點間的移動速度為實際交易資料移動速度
- (3) 每一台運補車的容量為 20 台自行車
- (4) 運補車搬上或是搬下一台自行車的時間為 1 分鐘

模擬測試分成兩個部份，第一部份是以某日的真實交易資料當作輸入檔，以真實交易情況下模擬一整天的運補指派；第二部份將多日之真實交易資料混合，再隨機選取以建構一日之模擬交易與運補指派，並重複此隨機測試多次（譬如 100 次）再觀察其平均表現。

#### 5.4.1 真實交易資料測試

真實交易資料的測試方法為以每日真實交易資料，測試在此交易情況下的運補指派決策，將會比較群眾運補與卡車運補的效益，亦會比較預測的成效。

表 5.7 模擬測試範例結果

日期	運補	預測	成功交易人次	失敗交易人次	召募人數	失敗率(%)
20140306	群眾	是	13536	502	3338	3.57
20140306	群眾	否	13513	504	2608	3.59
20140306	運補車	是	13247	759	0	5.41
20140306	運補車	否	13239	768	0	5.48
20140307	群眾	是	24202	707	3714	2.83
20140307	群眾	否	24112	777	3039	3.12
20140307	運補車	是	23726	1153	0	4.63
20140307	運補車	否	23687	1192	0	4.79
20140310	群眾	是	48174	2531	4492	4.99
20140310	群眾	否	47918	2791	3657	5.51
20140310	運補車	是	47171	3529	0	6.96
20140310	運補車	否	47230	3474	0	6.85
20140311	群眾	是	56852	3440	4628	5.71
20140311	群眾	否	56513	3751	3908	6.22
20140311	運補車	是	55794	4467	0	7.41
20140311	運補車	否	55386	4864	0	8.07

測試結果如表 5.7 所示，表 5.7 列出的是模擬了台北 YouBike 2014 年 3 月每日交易資料的範例結果，結果的欄位包含：

- (1) 運補：群眾運補或是運補車運補
- (2) 預測：此次模擬是否有加入預測考量
- (3) 成功交易人次：此次模擬成功完成借還需求的人次
- (4) 失敗交易人次：此次模擬借車失敗或是還車失敗的人次
- (5) 召募人數：召募多少群眾協助運補
- (6) 失敗率：失敗交易人次 ÷ 總交易人次

從表 5.7 結果範例可以看出，群眾運補的失敗率都比運補車運補來的低，如圖 5.8 將失敗率以折線圖表示（詳細數值參見表 A-1），平均一整天來說群眾運補的成功交易人次都比運補車運補多 2% – 3%。而群眾運補召募的可行性來說，平均召募的總人數為 4458 人，以台北 YouBike 共 164 個車站來看，平均每半小時在每個車站僅需要召募 0.566 個人即可達成每次的運補目標，此召募人數是一個合理且可接受的數字。

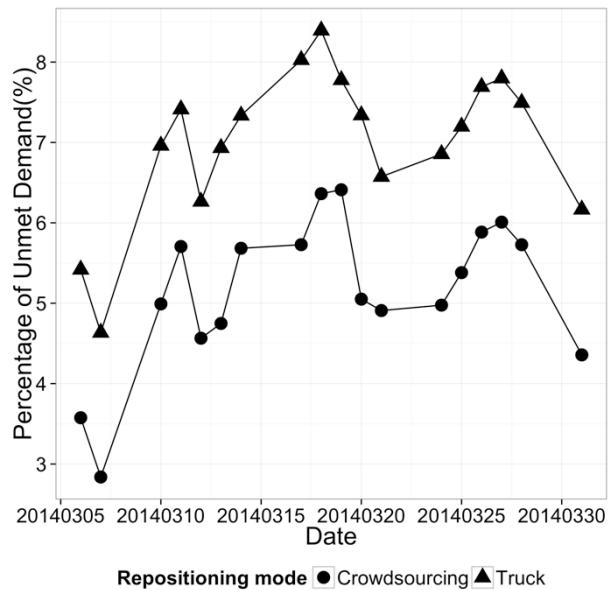


圖 5.8 群眾運補與運補車失敗率比較

圖 5.9 ( 詳細數值參見表 A-2 ) 顯示是否「加入預測機制」於群眾運補的失敗率比較結果，其中●為未使用預測機制、▲為使用預測機制，可以得知加入預測機制的確可有效減少群眾運補的失敗率。

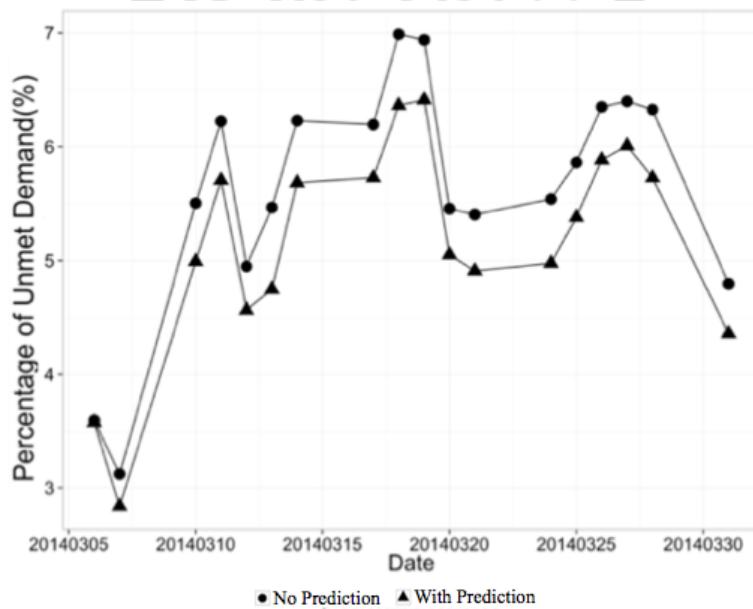


圖 5.9 群眾運補是否預測失敗率比較

上述的數值分析的卡車數皆設定為 10 台，本研究接著測試是否增加卡車的數量能有效的提昇成功交易人次。圖 5. 10 ( 詳細數值參見表 A- 3 ) 為將運補車數分別調整成 10、15、20、25 台的結果，可以發現當運補車多達 25 台時失敗率的確有下降，但其效益仍舊與群眾運補有段差距，其根本原因應在於 25 台卡車在任一時刻頂多僅能服務  $25/164=15\%$  的租借站，遠不如群眾運補方式可同步在所有租借站運補。

上述的分析結果皆僅看一整日累積下來的運補成效，算是一種平均效益。然而，在 PBSS 營運時，對使用者而言，他們可能更在意的是一天內各期( 特別是尖峰時期 ) 的失敗率，而非一整日的失敗率表現。因此我們將一天內各期的失敗率分別記錄比較，如圖 5. 11 ( 詳細數值參見表 A- 4 ) 所示，在尖峰時刻 8 點、9 點、18 點時，群眾運補的運補效益都明顯比卡車優，尤其在 18 點時，群眾運補改善約 32.5% 的失敗率。

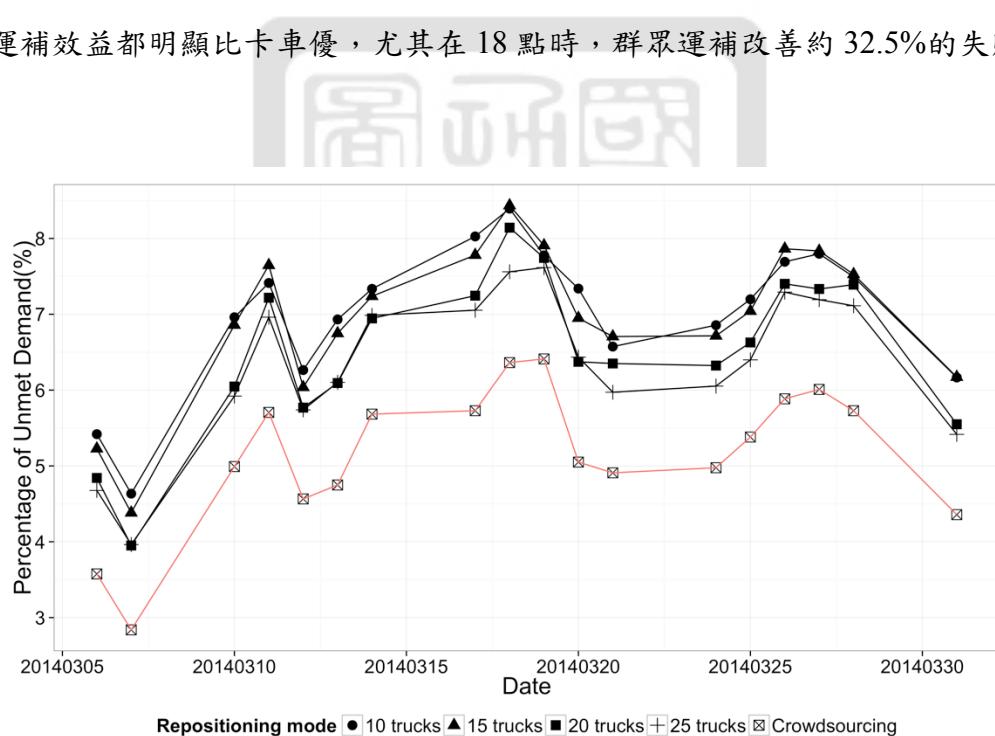


圖 5. 10 運補車數量與群眾運補失敗率比較

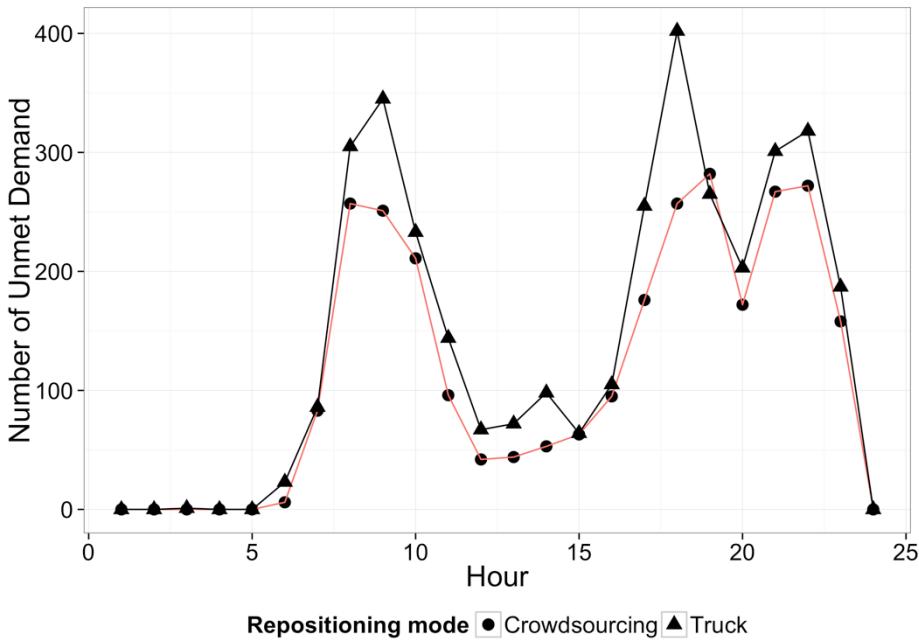


圖 5.11 運補車數量與群眾運補失敗率各時段比較

#### 5.4.2 隨機選取測試資料

在 5.4.1 節所測試的資料由於受限於真實交易資料的數量，所能測試的資料集不夠多，此外每日的真實交易狀況可能會因為當日可能發生的特殊情境（譬如慶祝活動、交通狀況、天氣等）而有特別的爆量或缺量，或許會因而影響其分析結果的一般性。為了避免這些特殊情境影響測試的分析結果，此測試將多日的租借資料混合在一起，再從中隨機挑選資料，並重複多次測試後再取其平均表現。舉例來說，若將某 3 天的交易資料混合在一起，每一筆交易資料會有  $1/3$  的機率會被選中，最後將選中的資料再混合作為測試資料。然而此種隨機選取測試中，由於每筆資料都有可能是來自不同天，因此當下的天氣參數、前期交易數等參數將因與其它被選取的資料獨立無關而變成不具參考性，因此以此方法測試就不採用隨機森林模型來協助預測，而是採用「歷史交易的平均值」作為預測值。舉例來說，從過去的資料統計出來，租借站 1 在 10:00-10:30 平均會有 10 個借車數，則這個 10 即會被用於估算該站在該期預期會發生的借車人次。

此測試做了一百次的隨機測試，並用台北 YouBike 2014 年 1-5 月的晴天作為隨機選取的資料集，測試結果如圖 5.12 所示，可以得知平均而言群眾運補的失敗率仍是比運補車運補來的低，亦即群眾運補的成效比運補車運補優。

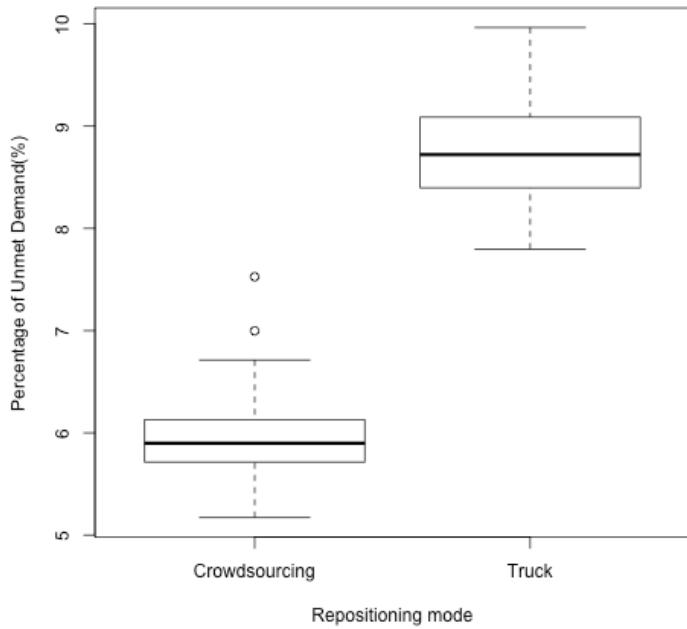


圖 5.12 群眾運補與運補車運補經過 100 次隨機選取資料測試後的失敗率比較

## 5.5 小結

本章的數值分析分成三個部份：第一部份為最適車數數學模式，可以依據不同的站點、日期、天氣狀況產生相對應的目標車數；第二部份是隨機森林模型，用以找出最適合台北 YouBike 的隨機森林模型。第三部份為模擬測試，用以比較群眾運補與運補車運補的效果。

第一部份的最適車數模型依據不同的參數產生各站點的目標車數，利用此車數可以幫助運補作業設定一個運補目標。第二部份的隨機森林模型，透過各站點以及不同參數的驗證，我們找到一個最適合台北 YouBike 的參數組合，並用此參數組合針對各

個站點產生各自的隨機森林模型，預先訓練好的模型可以加快模擬時的速度，減少每次都要重新產生模型的需要。

第三部份是模擬測試，實作群眾運補以及卡車運補的模擬程式，並且比較兩者效益。經過模擬測試的結果顯示，群眾運補在任何一天的運補結果都比卡車運補來的好，原因即為因為群眾運補可以將運補範圍同步擴展至所有站點，因此可以做更有效的運補。而增加運補車數雖然能夠提昇運補的成效，但其效益仍舊不及群眾運補方式。從數值分析結果可以得知有加入預測機制的運補會有效地減少失敗率，亦即需求預測能夠提昇運補的精確度。在群眾運補的可行性估算部分，從數值分析的結果可以看出平均每期每站號召不到一個人即可達成不錯的運補成效，這樣規模的號召人數理應可行，所以本研究認為群眾運補是一個可行且有效的方法。



## 第六章 結論與未來研究方向建議

### 6.1 結論與貢獻

公共自行車共享系統從 1960 年發展至今，全球已經有許多城市有自己的公共自行車共享系統，騎乘自行車對環境有益並具有健身效果，因此在各大都市都受到民眾的喜愛。然而在自行車共享系統一開始建制時，並非能完全的滿足當地民眾的使用習慣，因此容易發生熱門站的缺車或缺位現象，現行仰賴卡車協助平衡各租借站的需求，但此方法緩不濟急、花費的成本又多，因此本研究提出群眾運補的概念，期望利用群眾的力量協助運補，使得運補更有效益。本研究於第二章中，回顧了傳統 PBSS 運補及營運策略的文獻，了解過去對於靜態運補及動態運補的研究。為了將實際的交易結果與本研究結合，我們也回顧了一些分析實際公共自行車資料的相關研究。

過去的運補文獻大多假設一整天的交易需求為已知，再求解最佳的卡車路徑與卸載計劃，這些作法的缺點在於使用固定參數無法因應現實多變的需求，而且使用整數規劃模式求解卡車路徑與卸載計劃通常會耗費極大的運算時間，且僅能處理較小型網路，無法針對較大系統做即時的運補決，更遑論少數量的卡車無法同時服務多數量的租借站，因此卡車運補效益不彰是可以預期的。

本研究所提出的群眾運補方法可以處理大型共享系統之即時運補決策。我們以廖敏婷(2012)提出的人力自助運補概念為基礎，將一天切割成數期（譬如每 30 分鐘為一期），先假設隨時在各站皆可找到志願騎乘運補的人，計算出理論上各站在各期末的最佳自行車數量，以該數量當成各站各期使用即時運補的最佳目標車數；之後的即時運補則僅針對各站當下那期來規劃最佳之當期群眾運補方式，期以能使該站的期末自行車數量在該期歷經租還與運補之後，能盡量接近該最佳目標車數。為使決策能更精確，我們先以各站期初當下即時的自行車數為基礎，使用隨機森林模型預測當期內即將發生的借、還車量，計算各站在當期的理論應補或移車個數，再以貪婪法則優先

自應移走較多自行車的站，去運補其能力範圍內較需補進自行車的站，以此法則訂定兩兩站間的運補方式，直至該期期末再依實際租還量結算當下各站的自行車數，之後每期再重複這些步驟直至一天結束為止。本研究的運補機制有賴更精準的租借需求預測，因此必須先對該 PBSS 的歷史租借資料深度分析，以分析出的租借趨勢來協助預測的準確性，進而改進運補效益。

本研究在第三章針對 YouBike 2014 年 1-10 月的歷史資料做資料分析，首先了解所要處理的資料有哪些資料特性，我們將每一日的交易分別擷取出不同的資料屬性，搭配視覺化的工具輔助，從中分析台北 YouBike 以歸納出：(1)平日與假日的個別租借趨勢 (2)晴天與雨天的個別租借趨勢 (3)不同起訖點組合的相對比例趨勢。了解資料的特性後，我們將分析的結果用於第五章的模擬測試以及整數規劃模式，產出相關的數值分析結果。

第四章提出了本研究提出的兩個主要數學模式，第一個為最適車數數學模式，此模型的目的在於求出一個各站各期末的目標自行車數，如果在每期期末各站都能達到此目標值，則預期未來會發生缺車或缺位的可能性會降低。最適車數數學模式假設能夠號召無限多個志願騎乘者來幫忙運補，配合假設已知的各期各站租借需求量，可反推出每一期期末的自行車數，可將該數量視為各站當期期末最佳的自行車庫存量，此即為最適車數。第二個數學模式為當期群眾運補數學模式，為了處理當下即時運補的情況，僅考慮當下那一期的運補，試圖將各站的期初已知車數經由運補而達成該站期末的目標車數，如此每期期初重複這些步驟求解當期群眾運補數學模式，即可動態處理一整天的群眾運補。接著為了使運補指派更為有效，本研究利用隨機森林演算法做未來需求預測，帶入不同的參數尋找最適合的預測模型，本研究採用的是 Breiman (2001)提出的版本，並經過不同決策樹的數目、最低特徵值數量以及不同的特徵值選取找出最適合台北 YouBike 的預測模型，最後考量到運算的速度以及準確度，以 500 棵決策樹、3 個最低特徵值數量、特徵值組合「假日、工作日、降水量、溫度、前期

交易數、星期、小時」為最佳的隨機森林模型輸入參數值，利用此組合針對台北 YouBike 所有站點的借車、還車分別產出各站各期的個別隨機森林模型。

在第四章所算出來的各站各期之目標車數以及隨機森林預測模型將被用在第五章的模擬測試，以比群眾運補以及卡車運補的效益。從模擬的結果可以看出，群眾運補在所有測試中皆比卡車運補能更減少缺車或缺位的情況。此外，利用隨機森林的預測機制來輔助運補決策，的確會比未使用預測機制有更好的運補效益。

綜合以上所述，本研究目前貢獻如下：

1. **最適車數數學模式**：本研究提出的最適車數數學模式可以求出各站各期的理論最適車數，該車數可以提供現行系統營運管理參考。目前台北 YouBike 並沒有系統科學化的方式而僅以「經驗法則」來決定這個數值。最適車數數學模式以歷史租借資料的各期起訖需求平均值為基礎，求出各站於各期末的最佳目標自行車數。理論上，只要該站在該期期末的車數能被運補達此數值，則預期該站未來較不會發生缺車或缺位。此值可被當成各站各期的運補目標。
2. **隨機森林預測模型**：PBSS 的租借需求變動性極高，因此使用歷史租借資料的平均值來預測每站每期預期會發生的租還車數可能誤差過大，為了提高即時運補指派時的決策準確度，我們使用機器學習中的隨機森林演算法找出影響各站各期車數變化的最主要因子與參數組合，利用這些參數組合產生相對應的隨機森林預測模型，以估算當期即將發生的可能借還車量，則可以讓運補的指派更加精準，提昇運補效益。
3. **當期群眾運補數學模式**：當期群眾運補數學模式可視為最適車數數學模式的簡化版本，僅適用當期的即時運補指派，在已知當下期初各站車數、期末各站的最適車數、可招募到的群眾人數上限後，配合預測的本期租還車數（譬如使用隨機森林預測模型），即可以一整數規劃模式求解當期最佳運補方式。
4. **自行車共享系統分析視覺化**：目前全球有數百個 PBSS，目前大多數 PBSS 僅提供即時租借狀況網站，無法讓人更深入觀察 PBSS 的需求變化趨勢。本

研究依據分析台北 YouBike 的經驗，將這個數據分以及視覺化的過程包裝成一個程式流程，只要能提供 PBSS 的一些制式化租借資料（譬如租借站座標、容量、當下車數；每筆租借記錄的租還車時、地等），即可將該 PBSS 的一些相關深度分析的圖表以視覺化方式呈現，方便學習各 PBSS 的租借特性，並可比較不同 PBSS 的效益。

5. **群眾運補模擬流程**：為比較群眾運補與卡車運補的效益，本研究設計一模擬系統，以真實資料為租借資料來源，結合隨機森林預測模型，實作當期運補模式（可設定使用群眾運補或卡車運補）。為加速求解，我們提出一個貪婪演算法以處理當期的群眾運補模式，該法則優先針對互補效益較大的起訖點組合運補。由數據分析結果得知，群眾運補效益顯著優於卡車運補。
6. **可用於真實的運補指派**：過去的動態運補研究大都為紙上談兵，倚賴不夠準確的靜態租借資料平均值來求解整數規劃模式，因其規劃結果可能會與實際相差甚遠且耗時甚久，無法因應即時變動的租借需求。反之，本研究主要考慮動態的即時租借資料，而當期的運補模式求解容易且有效，同時使用機器學習中的隨機森林預測模型可提高參數的準確性，因此應該十分適合處理真實 PBSS 的運補指派決策。

## 6.2 建議之未來研究方向

本研究尚有不少未臻完善之處，以下列出幾點仍可延伸的研究議題：

1. **引發運補指派的時機**：本研究在「何時指派」運補，以每固定時段(每 30 分鐘)指派一次，不夠彈性，實際情況最好能使用「隨機應變」的方式，譬如當系統的運補需求累積至某種緊急程度時（譬如空車或空位少於某個門檻值）才引發群眾運補，然而該如何偵測判斷運補需求的「緊急程度」，是值得探討的議題。

2. **限制群眾運補的人數**: 在 4.3 節有提出群眾運補的理論數學模型，但因為受到測試資料時的限制，本研究僅用一個貪婪演算法求解，而非使用 Gurobi 求解該整數規劃模式，未來可以測試兩者的效益差別。
3. **需求分析的時段切割方式**: 本研究目前皆以每 30 分鐘作為時間切割的方式，時間點切割不同，會影響到每個時段所計算出來的交易數。理想的作法，可能要先針對歷史租借資料做更嚴謹的分析，針對每站的整日借車需求發生時刻，以更嚴謹的方式切時段，且每時段的長度不見得一定要相同。舉例來說，學理上經常會假設借車需求發生時刻遵守 Poisson Process ( 亦即 inter-arrival time 遵守 exponential distribution )，然而這個 Poisson Process 的期間究竟該如何訂定 ( 譬如是 06:00-07:00 ? 還是 06:16-07:48 ? 等時段切割問題 ) 才適當？甚至是否一定遵守 Poisson Process ? 若否，是否可找到更符合現實的機率分配方式？還是甚至根本某些時段的 inter-arrival time 本來就無規則可循？目前相關文獻幾乎沒看到類似的研究針對上述問題深入探討。若能更精準地切割時段，且各時段的需求 inter-arrival time 能找出更合理的理論機率分配，則可幫助預測需求的發生時刻，進而讓運補決策更正確有效。目前或可同時參考多種不同切段方式的分析結果，以淡化不同時間切割方式的影響。
4. **志願騎乘群眾的可獲性**: 本研究目前假設皆可找到群眾來運補的理想情況下，群眾運補的效果較佳，但實際運作時，可能必須配合有效的行銷手法 ( 譬如結合會員點數制，累積運補點數可兌換的紅利等等 )，才能找到足夠數量的志願騎乘者，未來或許可以試作可獲性的敏感度分析，看不同程度可獲性與目前運補車機制的效益比較。
5. **運補車及群眾運補結合**: 正常情況下，群眾運補應比卡車運補有更好的效益，然而在某些特例 ( 譬如下大雨或長距離上坡等 ) 的情況下，亦有可能難以找到願意幫忙騎乘的群眾，此時可能使用卡車運補較有效益，因此未來可以考慮將卡車及群眾運補結合，綜合規範不同運補方式的時機與範圍。

6. **更深度分析歷史租借資料集**：目前本研究有初步分析 10 個月的 YouBike 交易資料，但其前 5 個月仍為前 30 分鐘採免費的租借方式，因此使用者的使用習慣可能不會很一致。應盡量再分析更長時間（譬如 2、3 年）的租借資料，發掘與驗證是否各站各時段的租還需求存在某些規律，對資料更深度的理解將有益更精確的預測，也才能做出更佳的運補決策。
7. **運補車運補分群方式**：為比較運補車及群眾運補的效益，我們必須同時實作運補車的模擬機制。現狀採用先將租借站依運補車數分群，再各群自成一區以一車運補，然而分群原則如何決定並沒有一定的規則與定論。我們目前先僅使用 K-means 演算法來分群測試，因分群結果會連帶影響運補效益，未來可考慮測試不同分群方式對運補效益的差別。
8. **共享汽車及共享機車運補**：群眾運補這個概念可以延伸至共享汽車及共享汽車的運補上，因為就算使用卡車還是不方便搬運汽機車，感覺上似乎只能靠本研究提出的群眾運補方式。

## 參考文獻

- 周佰賢. (2015) 考慮需求變化狀況及增設臨停區之公共自行車共享系統租借站分群與車輛調度策略研究. 工業與資訊管理學系碩士論文，國立成功大學
- 洪菁蓬. (2011) 公共自行車共享系統之最佳租借站位址設置及車輛運補策略之研究. 工業與資訊管理學系碩士論文，國立成功大學
- 張立蓁. (2010) 都會區公共自行車共享系統之設計與營運方式研究. 工業與資訊管理學系碩士論文，國立成功大學
- 廖敏婷. (2012) 考慮需求比例及暫時人力配置之公共自行車共享系統管理策略研究. 工業與資訊管理學系碩士論文，國立成功大學
- Bargar, A., Gupta, A., Gupta, S., and Ma, D. "Interactive Visual Analytics for Multi-City Bike share Data Analysis." *The 3rd International Workshop on Urban Computing (UrbComp 2014)*, New York, USA. (2014).
- Breiman, L. "Random forests." *Machine learning* 45.1 (2001): 5-32.
- Cagliero, L., Cerquitelli, T., Chiusano, S., Garza, P., and Xiao, X. "Predicting critical conditions in bicycle sharing systems." *Computing*, (2016): 1-19.
- Contardo, C., Morency, C., and Rousseau, L. M. "Balancing a dynamic public bike-sharing system." *Centre Interuniversitaire de Recherche sur les Réseaux d'Entreprise, la Logistique et le Transport*, Vol. 4. (2012).
- Conway , M. W. "Predicting the Popularity of Bicycle Sharing Stations: An Accessibility-Based Approach Using Linear Regression and Random Forests." (2014).
- DeMaio, P. "Bike-sharing: History, impacts, models of provision, and future." *Journal of Public Transportation* 12.4 (2009): 3.

- Froehlich, J., Neumann, J., and Oliver, N. "Sensing and Predicting the Pulse of the City through Shared Bicycling." *International Joint Conference on Artificial Intelligence*. Vol. 9. (2009): 1420-1426.
- Gebhart, K., and Noland, R. B. "The impact of weather conditions on bikeshare trips in Washington, DC." *Transportation* 41.6 (2014): 1205-1225.
- Kaltenbrunner, A., Meza, R., Grivolla, J., Codina, J., and Banchs, R. "Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system". *Pervasive and Mobile Computing*, 6(4), (2010): 455-466.
- Kaspi, M., Raviv, T., Tzur, M., and Galili, H. "Regulating vehicle sharing systems through parking reservation policies: Analysis and performance bounds." *European Journal of Operational Research*, 251(3), (2016): 969-987.
- Montoliu, R. "Discovering mobility patterns on bicycle-based public transportation system by using probabilistic topic models." *Ambient Intelligence-Software and Applications*. Springer Berlin Heidelberg, (2012): 145-153.
- O'brien, O., Cheshire, J., and Batty, M. "Mining bicycle sharing data for generating insights into sustainable transport systems." *Journal of Transport Geography* 34 (2014): 262-273.
- O'Mahony, E. D. "Smarter Tools For (Citi) Bike Sharing. " PhD Thesis. Cornell University. (2015).
- Raviv, T., Tzur, M., and Forma, I. A. "Static repositioning in a bike-sharing system: models and solution approaches." *EIRO Journal on Transportation and Logistics* 2.3 (2013): 187-229.
- Rixey, R. "Station-Level Forecasting of Bikesharing Ridership: Station Network Effects in Three US Systems." *Transportation Research Record: Journal of the Transportation Research Board* 2387 (2013): 46-55.

- Sarkar, A., Lathia, N., and Mascolo, C. "Comparing cities' cycling patterns using online shared bicycle maps." *Transportation* 42.4 (2015): 541-559.
- Vogel, P., and Mattfeld, D. C. "Strategic and operational planning of bike-sharing systems by data mining—a case study." *International Conference on Computational Logistics. Springer Berlin Heidelberg.* (2011)
- Vogel, P., Greiser, T., and Mattfeld, D. C. "Understanding bike-sharing systems using data mining: Exploring activity patterns." *Procedia-Social and Behavioral Sciences* 20 (2011): 514-523.
- Vogel, P., Saavedra, B. A. N., and Mattfeld, D. C. "A hybrid metaheuristic to solve the resource allocation problem in bike sharing systems. " *International Workshop on Hybrid Metaheuristics. Springer International Publishing.* (2014): 16-29.
- Yang, Z., Hu, J., Shu, Y., Cheng, P., Chen, J., and Moscibroda, T. "Mobility Modeling and Prediction in Bike-Sharing Systems". *MobiSys '16 Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services* (2016): 165-178
- Yoon, J. W., Pinelli, F., and Calabrese, F. "Cityride: a predictive bike sharing journey advisor." *2012 IEEE 13th International Conference on Mobile Data Management. IEEE,* (2012): 306-311

## 附錄 A、模擬數值結果

表 A-1 圖 5.8 數值

日期	群眾運補(%)	卡車運補(%)
20140306	3.58	5.42
20140307	2.84	4.63
20140310	4.99	6.96
20140311	5.71	7.41
20140312	4.56	6.26
20140313	4.75	6.93
20140314	5.68	7.34
20140317	5.73	8.03
20140318	6.36	8.39
20140319	6.41	7.78
20140320	5.05	7.34
20140321	4.91	6.57
20140324	4.98	6.86
20140325	5.38	7.2
20140326	5.88	7.69
20140327	6.01	7.8
20140328	5.73	7.49
20140331	4.36	6.17

表 A-2 圖 5.9 數值

日期	有預測(%)	無預測(%)
20140306	3.58	3.6
20140307	2.84	3.12
20140310	4.99	5.5
20140311	5.71	6.22
20140312	4.56	4.95
20140313	4.75	5.47
20140314	5.68	6.23
20140317	5.73	6.2
20140318	6.36	6.99
20140319	6.41	6.94
20140320	5.05	5.45
20140321	4.91	5.4
20140324	4.98	5.54
20140325	5.38	5.86
20140326	5.88	6.35
20140327	6.01	6.4
20140328	5.73	6.33
20140331	4.36	4.8

表 A-3 圖 5.10 數值

日期	群眾運補 (%)	10 台運補車 (%)	15 台運補車 (%)	20 台運補車 (%)	25 台運補車 (%)
20140306	3.58	5.48	5.28	4.6	4.5
20140307	2.84	4.79	4.62	4.21	4.3
20140310	4.99	6.85	6.65	6.63	6.66
20140311	5.71	8.07	8.01	7.59	7.66
20140312	4.56	6.55	6.23	6.18	6.01
20140313	4.75	7.04	7	6.64	6.55
20140314	5.68	7.51	7.15	7.07	7.09
20140317	5.73	8.15	8.23	7.59	7.2
20140318	6.36	8.97	8.89	8.23	8.1
20140319	6.41	8.02	7.99	7.75	7.71
20140320	5.05	7.75	7.05	6.64	6.74
20140321	4.91	6.74	6.83	6.5	6.23
20140324	4.98	6.73	6.71	6.76	6.37
20140325	5.38	7.51	7.03	7.03	6.73
20140326	5.88	7.99	7.93	7.7	7.71
20140327	6.01	7.98	7.76	7.54	7.59
20140328	5.73	7.79	8.13	7.37	7.29
20140331	4.36	6.17	5.97	5.75	5.87

表 A-4 圖 5.11 數值

時段	群眾運補	卡車運補
0	0	0
1	0	0
2	0	1
3	0	0
4	0	0
5	6	23
6	83	86
7	257	305
8	251	345
9	211	233
10	96	144
11	42	67
12	44	72
13	53	98
14	63	64
15	95	105
16	176	255
17	257	402
18	282	265
19	172	203
20	267	301
21	272	318
22	158	187
23	0	0