

逢甲大學學生報告 ePaper

機器學習應用於共用單車流量分佈

Machine learning applied to shared bicycle traffic distribution

作者：陳澤天

系級：建築五甲

學號：D0261705

開課老師：陳上元

課程名稱：建築設計(十)

開課系所：建築專業學院

開課學年：106 學年度 第 2 學期

中文摘要

隨著共用單車在城市生活中大規模普及和應用，不同區塊之間的供需失衡問題日益嚴重，開始影響市容和居民的正常出行。因此共用單車的及時、或者是超前調度顯得尤為重要。本研究建議應用機器學習（Machine Learning）聚類分析（Cluster Analysis）於交通類工具的流量分佈，它根據在地化的數據資訊，進行模擬演算，產生即時的視覺化流量資訊，同時分辨有任意形狀或者密度特性的資訊類別。除此之外，無樁模式的共用單車具有取車和還車地點的隨機性，如何架構合理的單車存量分區方式、計算分區內的單車密度、並根據單車密度作為參數控制的條件、以提出相對應的單車疏散與調度策略成為本研究的重點。

本研究採用 R 語言作為處理和分析數據、進行劃分區塊操作的程式語言。以虛擬環境軟體 Rstudio，作為進行聚類分析和資料處理的載體。運用機器學習聚類分析有三種方法，包括：K-平均演算分群法（K-means）、密度分群法以及 DBSCAN 分群法（Density-based spatial clustering of applications with noise）。

本文以上海市靜安寺周邊以及上海四環（ 121.75°E - 121.125°E ， 30.95°N - 31.45°N ）為研究範圍，探討兩種範圍內摩拜單車在單位時段內的取得量和歸還量，借鑒紐約市計程車與城市單車在聚類分析下的分區方式，探索最佳的上海共用單車分區方式。在最佳分區方式下，本文以預測相鄰時段內單車騎行目的地區塊的趨勢分佈為基礎，提出基於閾值控制思維的逆向流動獎勵制度作為單車的疏散與調度策略。

機器學習應用於共用單車流量分佈

實作發現，在僅考量不同時間段內單車的流量變化時，以上海市靜安寺及其周邊為尺度的研究範圍的情況下，應採用用密度分群法，根據取車點和還車點的分佈建立各自獨立的獎勵機制分區，分別給予獎勵，得到的單車流動趨向較為合理，會使實施後的結果趨向于平衡狀態；以上海四環為尺度的研究範圍，應採用 DBSCAN 分群法，根據還車點的分佈建立密度視覺化監督系統，使得運營決策者能夠及時瞭解單車在不同密度值下呈現的分佈狀態，並以此做參考，來決定下一階段新投入運營單車的數量和投放地點。

未來，隨著資料透明化的推行，精確度更高的資料可以公開使用到學術領域，大量資料的導入訓練，可增強本研究結果的分區準確度和更短週期的獎勵方案，使獎勵機制、調度策略更加具體。

關鍵字：機器學習、聚類分析、供需平衡、趨勢分佈、調度機制

Abstract

With the widespread adoption and application of shared bicycles in urban life, the problem of supply and demand imbalances between different blocks has become increasingly serious, and it has begun to affect the normal appearance of city appearance and residents. Therefore, timely or advanced scheduling of shared bicycles is particularly important. This study proposes the application of Machine Learning Cluster Analysis to the traffic distribution of traffic tools. Based on the localized data information, it performs simulation calculations to generate instant visual traffic information and distinguishes any Information categories for shape or density characteristics. In addition, the shared bicycle without pile mode has the randomness of the pick-up and drop-off locations, how to structure a reasonable method of partitioning the single-vehicle inventory, calculate the density of the bicycle in the partition, and use the density of the bicycle as a parameter control condition to propose Corresponding bicycle evacuation and scheduling strategies have become the focus of this study.

This study uses R language as a programming language for processing and analyzing data and dividing blocks. The virtual environment software Rstudio is used as a carrier for cluster analysis and data processing. There are three methods for cluster analysis using machine learning, including: K-means, density clustering, and DBSCAN grouping (Density-based spatial clustering of applications with noise). In this paper, around the Jing'an Temple in Shanghai and Shanghai Sihuan (121.75°E-121.125°E, 30.95°N-31.45°N) as the research scope, the acquisition and return of Mobike bicycles in two ranges within a unit time are discussed. Invest in the zoning method of the New York City Taxi under cluster analysis to explore the best Shanghai shared bike partition method. Under the optimal partition method, based on

the prediction of the trend distribution of bicycle riding destination blocks in adjacent periods, a reverse flow reward system based on threshold control thinking was proposed as the evacuation and scheduling strategy of bicycles.

It is found that in the case of considering only the flow of bicycles in different time periods and using the scale of the Jing'an Temple and its surroundings as the scope of the study, a density clustering method should be used, based on the point of pick-up and return of the vehicle. Distribution and establishment of independent incentive mechanism zoning, respectively rewarded, the resulting bicycle flow tends to be more reasonable, will result in the results after the implementation tends to a balanced state; the scope of the study in Shanghai Sihuan as the yardstick, the DBSCAN grouping method should be used, according to The distribution of vehicle points establishes a visual density monitoring system, which enables the operational decision makers to know the distribution status of bicycles at different density values in a timely manner, and uses this as a reference to determine the number of new bicycles to be put into operation in the next phase and the location of the new bicycles. In the future, with the implementation of transparent data, more accurate data can be publicly used in the academic field. The introduction of a large amount of data can enhance the partition accuracy of this research result and a shorter-cycle reward program, enabling the reward mechanism. The scheduling strategy is more specific.

Keyword : Machine Learning 、 Cluster Analysis 、 Supply-demand Balance 、 Trend Distribution 、 Scheduling Mechanism

目次

摘要.....	錯誤! 尚未定義書籤。
Abstract.....	錯誤! 尚未定義書籤。
目錄.....	錯誤! 尚未定義書籤。
圖目錄.....	7
表目錄.....	9
第一章 緒論.....	10
1.1.....	研究動機與目的
.....	11
1.2.....	研究範圍與內容
.....	13
1.3.....	研究方法與流程
.....	15
1.4.....	研究課題與限制
.....	17
1.5.....	用語定義
.....	19
第二章 文獻回顧.....	23
2.1.....	共用單車發展史
.....	25
2.1.1.....	共用單車的起源
.....	25
2.1.2.....	共用單車的發展
.....	25
2.2.....	無樁共用單車使用流程
.....	28
2.2.1.....	第一代無樁共用單車
.....	28
2.2.2.....	第二代無樁共用單車
.....	29
2.3.....	單車流量之統計與預測
.....	31
2.3.1.....	有樁單車流量統計與預測之分區方式
.....	31
2.3.2.....	無樁共用單車流量統計與預測之分區方式
.....	37
2.4.....	聚類分析及其分支下三種分群法

.....	38
2.4.1	K-平均演算分群法
.....	38
2.4.2	密度分群法
.....	39
2.4.3	DBSCAN 分群法
.....	40
2.5.....	小結
.....	43
第三章 理論與方法.....	45
3.1.....	基礎工具和程式語言
.....	46
3.2.....	基於閾值控制思維的逆向流動獎勵制度及其沙盤推演
.....	47
3.2.1	基於閾值控制思維的逆向流動獎勵制度
.....	47
3.2.2	獎勵機制的沙盤推演
.....	48
3.3.....	小結
.....	54
第四章 實作分析.....	55
4.1.....	階段一：調取並處理數據
.....	57
4.2.....	階段二：探究變量對於預測系統的影響
.....	63
4.2.1	計程車上客點位分析
.....	63
4.2.2	單車數據資料精確度
.....	64
4.3 階段三：分群方法的探究及其實際運用.....	83
4.3.1 應用密度分群法，處理較小的街道尺度，以上海靜安寺及其周邊為例.....	83
4.3.2 應用 DBSCAN 分群法，處理較大的城市尺度，以上海四環及其內部為例.....	90
4.4 階段四：疏散，調度策略總結.....	100
4.5 小結.....	101
第五章 討論與建議.....	102
5.1 調度策略總結.....	103

5.2 後續研究建議.....	104
參考文獻.....	錯誤! 尚未定義書籤。
中文.....	105
英文.....	105
網頁.....	106

圖目錄

圖 1-1 四環 (121.75°E-121.125°E, 30.95°N-31.45°N) 範圍單車還車點分佈.....	14
圖 2-1 研究流程.....	16
圖 2-1 第一代公用單車使用流程.....	28
圖 2-2 第二代公用單車使用流程.....	29
圖 2-3 群組間單車流動機率.....	31
圖 2-4 紐約市城市單車分區示意圖.....	32
圖 2-5 紐約城市單車分區系統架構簡圖.....	34
圖 2-6 K-平均演算分群法演示.....	37
圖 2-7 核密度估算的視覺化.....	38
圖 2-8 核密度估算.....	38
圖 2-9 DBSCAN 分群法原理.....	39
圖 2-10 區域劃分分區示意圖.....	40
圖 3-1 Rstudio 的使用者介面.....	43
圖 3-2 共用單車控制系統模型.....	45
圖 3-3 沙盤分區示意圖.....	46
圖 3-4 分區的單車單位元元密度.....	46
圖 3-5 分區的單車還車速度 (check out)(左)與取車速度 (check in)(右)	47
圖 3-6 獎勵機制示意.....	48
圖 3-7 總獎勵額度計算.....	49
圖 4-1 上海市 2016 年 8 月 1 日 (禮拜一) 00:00 至 24:00 取車點示意圖.....	55
圖 4-2 上海市 2016 年 8 月 1 日 (禮拜一) 00:00 至 06:00 取車點示意圖.....	55
圖 4-3 上海市 2016 年 8 月 1 日 (禮拜一) 06:00 至 12:00 取車點示意圖.....	56
圖 4-4 上海市 2016 年 8 月 1 日 (禮拜一) 12:00 至 18:00 取車點示意圖.....	56
圖 4-5 上海市 2016 年 8 月 1 日 (禮拜一) 18:00 至 24:00 取車點示意圖.....	57
圖 4-6 上海市 2016 年 8 月 4.5 日 (禮拜四、五) 取車點示意圖.....	57
圖 4-7 上海市 2016 年 8 月 6.7 日 (禮拜六、日) 06:00 至 12:00 取車點示意圖.....	58
圖 4-8 上海市 2016 年 8 月 4.5 日 (禮拜四、五) 00:06 至 00:12 取車點熱力示意圖.....	58
圖 4-9 上海市 2016 年 8 月 6.7 日 (禮拜六、日) 00:06 至 00:12 取車點熱力示意圖.....	59

機器學習應用於共用單車流量分佈

- 圖 4-10 紐約市計程車上客點分佈圖……61
- 圖 4-11 紐約市計程車上客點等高線圖……62
- 圖 4-12 上海摩拜單車還車點分佈圖 a……63
- 圖 4-13 上海摩拜單車還車點分佈圖 b……64
- 圖 4-14 上海摩拜單車還車點分佈圖 c……65
- 圖 4-15 上海摩拜單車還車點分佈圖 d……66
- 圖 4-16 上海摩拜單車還車密度曲線分佈圖 a……67
- 圖 4-17 上海摩拜單車還車密度曲線分佈圖 b……68
- 圖 4-18 上海摩拜單車還車密度曲線分佈圖 c……69
- 圖 4-19 上海摩拜單車還車密度曲線分佈圖 d……70
- 圖 4-20 上海摩拜單車還車密度曲線分佈圖……71
- 圖 4-21 上海摩拜單車取車密度曲線分佈圖……72
- 圖 4-22 上海楊浦區摩拜單車還車密度曲線分佈圖……73
- 圖 4-23 上海楊浦區摩拜單車取車密度曲線分佈圖……74
- 圖 4-24 上海五角場摩拜單車還車密度曲線分佈圖……75
- 圖 4-25 上海五角場摩拜單車取車密度曲線分佈圖……76
- 圖 4-26 上海五角場地鐵站摩拜單車還車密度曲線分佈圖……77
- 圖 4-27 上海五角場地鐵站摩拜單車取車密度曲線分佈圖……78
- 圖 4-28 上海靜安寺摩拜單車還車密度曲線分佈圖……79
- 圖 4-29 上海靜安寺摩拜單車取車密度曲線分佈圖……80
- 圖 4-30 550m *550m 範圍點位示意圖……81
- 圖 4-31 根據點位元上存在的點密度分群示意圖……82
- 圖 4-32 上海靜安寺 2016.8 還車密度分區……85
- 圖 4-33 上海靜安寺 2016.8 取車密度分區……86
- 圖 4-34 單車於還車分區獎勵示意圖……88
- 圖 4-35 單車於取車分區獎勵示意圖……89
- 圖 4-36 K-平均演算分群法對於紐約計程車上客點位分析……91
- 圖 4-37 K-平均演算分群法對於紐約計程車上客點位地圖……92
- 圖 4-38 參數設定相同 K-平均演算分群法對於紐約計程車上客點位地圖……95
- 圖 4-39 DBSCAN 分群法對於紐約計程車上客點位分析……94
- 圖 4-40 DBSCAN 分群法對於紐約計程車上客點位地圖……95
- 圖 4-41 掃描半徑 700m 範圍示意圖……96
- 圖 4-42 還車點位，每個點位上有大於 32 個點單位元面積的密度標示……97
- 圖 4-43 還車點位，每個點位上有大於 57 個點單位元面積的密度標示……98
- 圖 4-44 還車點位，每個點位上有大於 124 個點單位元面積的密度標示……98

表目錄

表 2-1 紐約市城市單車還車分區誤差率表……32

表 2-2 紐約市城市單車取車分區誤差率表……33

表 4-1 還車分區各級數區域閾值……87

表 4-2 取車分區各級數區域閾值……88

第一章 緒論

1.1 研究動機與目的

1.2 研究範圍與內容

1.3 研究方法與流程

1.4 研究課題與限制

1.5 用語定義

第一章 緒論

1.1 研究動機與目的

面對全球氣候變化與溫室氣體過量排放的情形，低碳出行可以有效的減緩交通所帶來的空氣污染和溫室效應 (Greenhouse effect)，許多國家和政府紛紛設立資金會支援低碳交通出行方式的開發和相關項目的建立，以提高城市可持續發展 (Sustainable Development) 的能力，創造環境友善 (Environment-friendly) 的都市面貌。

未來，城市交通必定要求民眾選擇更多的低碳環保出行方式，其中面對的兩大挑戰是：(一) 如何滿足隨著城市人口指數增長而增長的交通需求、(二) 如何建立更完善的低碳交通系統覆蓋民眾各種距離的出行需求。

一、滿足隨著城市人口指數增長而增長的交通需求，旨在建立能獨立承擔起交通運輸功能的項目。出行具有巨大的市場容量，這同時是挑戰也是機遇，對於企業，意味著優秀的運營可以帶來非常豐厚的後期回報。二、建立更完善的低碳交通系統覆蓋民眾各種距離的出行需求，旨在發掘短途出行的價值。在中國大陸市場，Uber 和滴滴打車之前，沒有任何機構有資料來支撐短途出行這個需求的價值空間。只有計程車公司那裡有一些碎片化的資料。而 Uber 和滴滴打車的營收，顯示出出行領域最大的市場，就是短途出行。3-5公里的短途出行，佔據兩大公司營收的七成以上。

在此根據上，共用單車的概念在中國大陸一進出現，就收到了熱烈的追捧。2016年在各個大型城市中普及後，引來了許多資本的多輪追加融資。伴隨著資本的湧入，共用單車的運營模式也逐漸開始發展和創新。

資本湧入的同時，運營策略的更新卻落後於其發展的態勢。尤其出現了不同區塊間供需失衡的問題。在某些重要的交通節點，比如地鐵站、客運中心、景點等等，高峰期間經常出現單車數量囤積、或者是數量匱乏的現象。單車公司運營者往往採取事態後調度的方式來減輕運營的壓力。然而，一般根據單車囤積、或者匱乏現象產生後才進行調度的方式，往往緩不濟急。單車公司需要調派人力、和運輸車輛把囤積區域的單車運送至匱乏區域，產生了物流成本的損耗。在此同時，處於運輸狀態的單車，也意味著本身屬於不可被使用的狀態，因此降低了城市交通的承載能力，所以共用單車的及時、或者是超前調度顯得尤為重要。

同時，在大數據環境下，資料科學（data science）逐漸肩負起分析巨量資料的重任，許多原本用傳統方法採集和分析資料的方式逐漸被人工智慧（artificial intelligence）所取代。在交通領域，例如地鐵、公交、計程車或公共單車，出現全面的開放資料（open data），可以讓使用者從中解析出真實而全面的交通規律。本研究藉由交通資料公開化所帶來大量真實資料的契機，尋找並分析公開、合法的無樁共用單車資料，預測無樁共用單車可能囤積或者匱乏現象發生的時間和區塊，以協助單車公司決策者能夠提前規劃調度，以提升城市交通的承載能力、和降低調度不良所產生的風險與成本。

1.2 研究範圍與內容

本研究藉由資料科學的相關技術，通過對紐約市計程車資料 (NYCopen data, 2018)，紐約市城市單車資料 (CitiBike, 2018)，上海市摩拜單車資料 (Sodaopendata, 2018) 等相關交通資料的整理和分析，探究交通類資料在實際區域上的分佈。

同時，研究應用機器學習 (Machine Learning) 聚類分析 (Cluster Analysis) 於交通類工具的流量分佈，根據在地化的數據資訊，進行模擬演算，產生即時的視覺化流量資訊，從而分辨有任意形狀或者密度特性的資訊類別。除此之外，無樁模式的共用單車具有取車和還車地點的隨機性，如何架構合理的單車存量分區方式、計算分區內的單車密度、並根據單車密度作為參數控制的條件、以提出相對應的單車疏散與調度策略成為本研究的重點。

研究以上海市靜安寺周邊以及上海四環 (121.75°E - 121.125°E ， 30.95°N - 31.45°N) 為研究範圍，探討兩種範圍內摩拜單車在單位時段內的取得量和歸還量，借鑒紐約市計程車在聚類分析下的分區方式，探索最佳的上海共用單車分區方式。在最佳分區方式下，本文以預測相鄰時段內單車騎行目的地區域的趨勢分佈為基礎，提出基於閾值控制思維的逆向流動獎勵制度作為單車的疏散與調度策略。研究範圍的選擇基於表現兩種不同尺度下分群方法的應用，街道尺度選取上海老城區中，屬於商業、住宅中心的靜安寺區域作為研究對象；城市尺

機器學習應用於共用單車流量分佈

度選取上海摩拜單車主要運行覆蓋區域，四環公路及其以內作為研究對象，如圖 1-1。

本研究採用 R 語言作為處理和分析數據、進行劃分區塊操作的程式語言。以虛擬環境軟體 Rstudio，作為進行聚類分析和資料處理的載體。同時，在探究最佳分區的過程中，運用機器學習聚類分析 (Cluster Analysis) 的方法，探討何種方法和理論最為適用於實際場景。

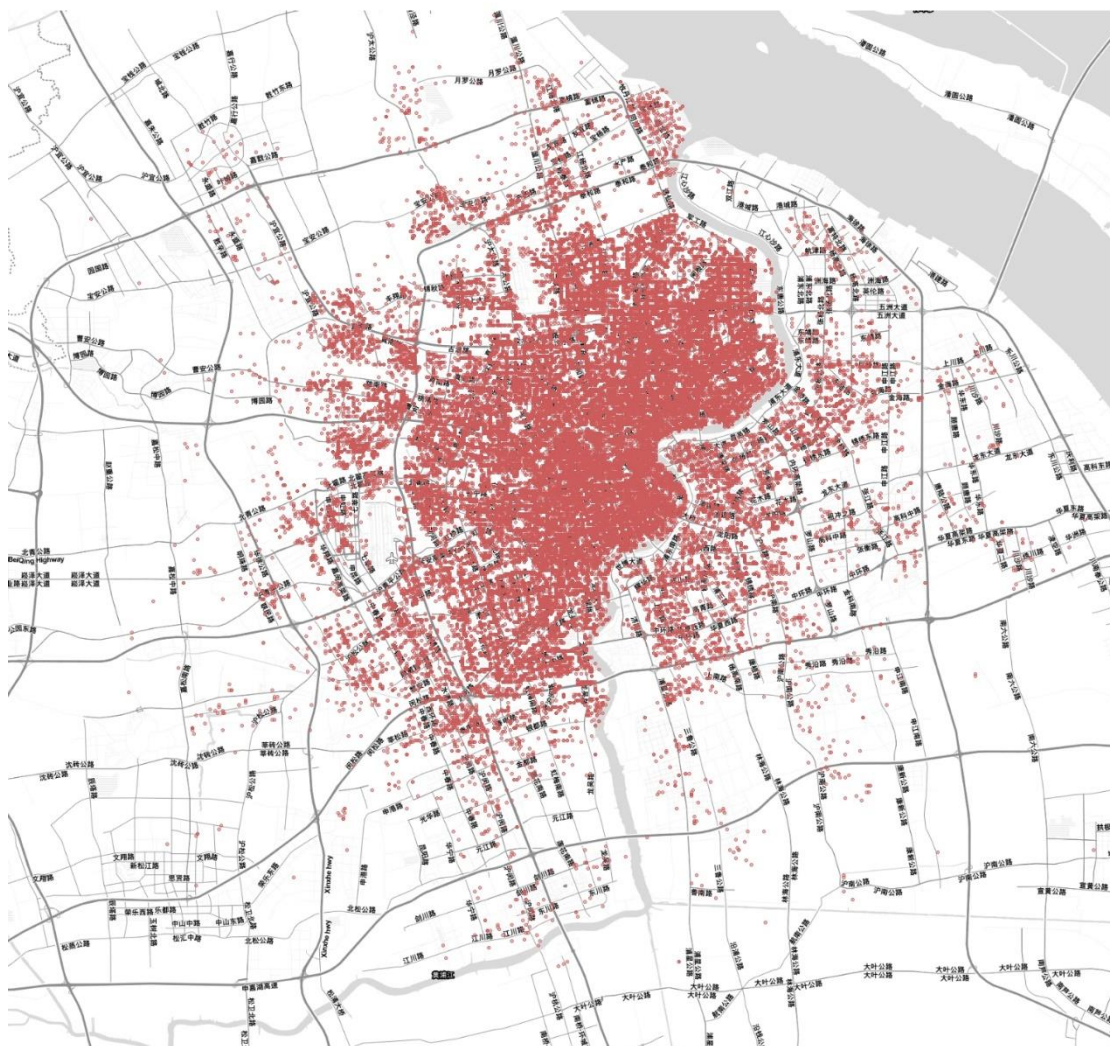


圖 1-1 四環 (121.75°E-121.125°E, 30.95°N-31.45°N) 範圍單車還車點分佈

1.3 研究方法與流程

根據研究範圍與內容的界定，本研究計劃步驟為緒論、文獻回顧、理論與方法、實作分析、討論與建議等五個部分，說明如下：

緒論：釐清研究動機與目的，確定研究範圍與內容，發展研究方法與流程，建立研究課題，整理與理解用語和相關定義。

文獻回顧：本文研究的對象為無樁模式的共用單車，無樁模式的共用單車具有取車和還車地點的隨機性，因此如何架構合理的單車存量分區方式、計算分區內的單車密度、並根據單車密度作為參數控制的條件、以提出相對應的單車疏散、與調度策略成為本研究的重點。本章就共用單車發展史、無樁共用單車使用流程、以及單車流量之統計與預測三個方面的文獻與案例，收集本研究發展所需之條件。

理論與方法：本章節就模擬化的獎勵機制及其沙盤推演為基礎，闡述實際運用獎勵機制時要面對的邏輯問題，及對應的策略與方案。分為（一）基礎工具和程式語言、（二）基於閾值控制思維的逆向流動獎勵制度及其沙盤推演。

實作分析：在實際測試運用，分析比較的過程中，分別以靜安寺及其周邊、上海四環（121.75°E-121.125°E，30.95°N-31.45°N）及其內部為兩種不同空間尺度的研究範圍，探討在不同空間尺度下，最佳的分區方式。

討論與建議：實際運用時應結合所面對的現象的具體情況，根據不同分群方法的優缺點，採用合理的分群方法處理資料。

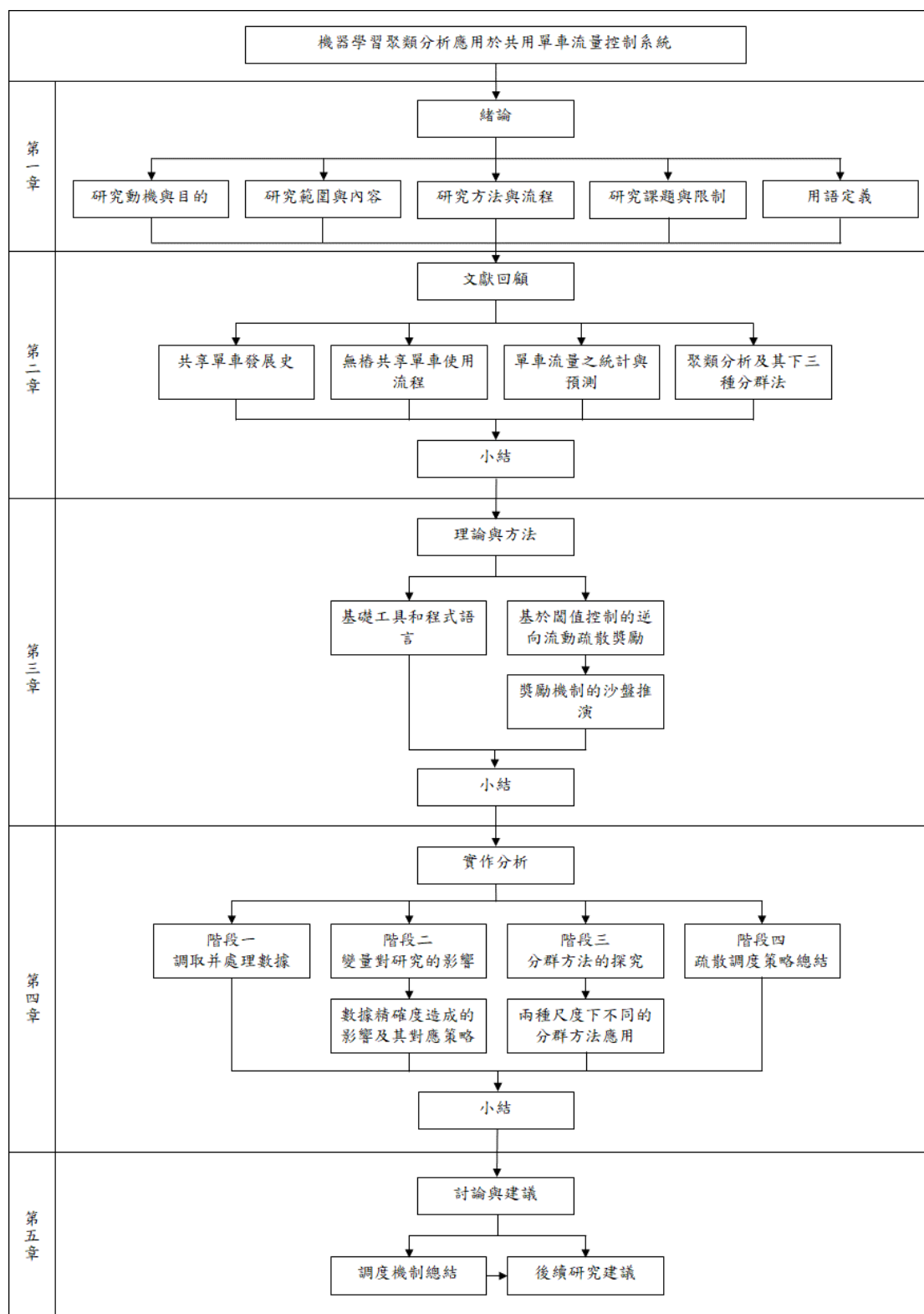


圖 2-1 研究流程

1.4 研究課題與限制

(1) 研究課題

單車業者具備及時、或者是超前調度的管理能力十分重要。針對基於邏輯理論下的基本分區方式，建立合理的獎勵機制。根據不同實際情況，探討適合不同情形下的聚類學習分區方式，並建立實際的獎勵機制，同時通過反向邏輯推理的方式驗證獎勵或者調度機制的合理性、可行性。

▲基於閾值控制思維的逆向疏散獎勵機制

此階段進行對於閾值控制思維的闡述，根據其原理，以及基礎的分區方式擬定獎勵機制，並進行沙盤推演。

▲在實際應用中探究三種聚類分析方法及其合適使用情境

此階段根據三種聚類分析方法的原理與特性，實際應用於計程車與城市無樁公用單車之中，通過對數據資料的視覺化，探究不同情況下點位的分佈，選擇合適的聚類分析方法，並建立實際的獎勵或調度機制。

(2) 研究限制

針對研究對象無樁式共用單車，該模式的共用單車由於發展時間有限，學術界並沒有很多合適的論文可以借鑒和參考。

對於無樁式共用單車，流通於世界上的項目多為私人企業、上市公司所開發與擁有，所以單車相關的資料或者數據都為私人、公司所有，公開可以流通的資料非常少，而且流通資料的精確度不高。根據此種限制，本研究中單車密度分區的建立、以及調度制度的設定均以現有資料為分析基礎，以每月或者每天的度量單位來研究單車密度變化的規律，以指代全面數據公開條件下的每分或每秒的度量單位。

同時，流通的資料具有滯後性，即一般不會有即時的數據資料被公開。根據此種限制，本研究中使用可獲得的歷史數據資料來研究相關課題。

作為在地化的交通工具，無樁共用單車的運動軌跡、規律會因為在地化的客觀因素導致和資料所呈現的現象有所不同。本研究暫不考量在地化因素如：政策，人文……等所帶來的對於研究的影響。

1.5 用語定義

在對於方法論探討、劃分集群、程式操作、策略闡述等過程中，有許多屬於學科內的專用術語，本節將這些詞彙逐一辨析，以釐清本研究所要研究的對象範圍。針對本研究內容，整理相關的用語定義，說明如下：

▲人工智慧 (Artificial intelligence)

是指由人製造出來的機器所表現出來的智慧。通常人工智慧是指通過普通電腦程式的手段實現的人類智慧技術。

人工智慧的研究是從以“推理”為重點到以“知識”為重點，再到以“學習”為重點。

▲機器學習 (Machine Learning)

機器學習是人工智慧一個分支。機器學習在近30多年已發展為一門多領域交叉學科，涉及概率論、統計學、逼近論、凸分析、計算複雜性理論等多門學科。

機器學習演算法是一類從數據中自動分析獲得規律，並利用規律對未知數據進行預測的演算法。機器學習可以分成下面幾種類別：監督學習 (Supervised learning)、無監督學習 (Unsupervised learning)、增強學習 (Reinforcement learning)。

▲聚類分析 (Cluster analysis)

機器學習應用於共用單車流量分佈

聚類分析是機器學習無監督學習分支下的一種技術，聚類是把相似的物件通過靜態分類的方法分成不同的組別或者更多的子集 (Subset)，這樣讓在同一個子集中的成員物件都有相似的一些屬性。

其中基於密度的聚類演算法，是為了挖掘有任意形狀特性的類別而發明的。此演算法把一個類別視為資料集中大於某閾值的一個區域。

▲K-平均演算法 (K-means)

聚類分析分支下的一種演算法

▲DBSCAN (Density-based spatial clustering of applications with noise)

聚類分析分支下的一種演算法

▲R 語言 (R programming)

多種程式語言 (programming language) 的其中一種，主要用於統計分析、繪圖、資料採擷。其他程式語言如 C 語言、Python 等。

▲集成開發環境 (Integrated Development Environment)

又稱整合開發環境，IDE，虛擬環境，虛擬器，是一種輔助程式開發人員開發軟體的應用軟體，在開發工具內部就可以輔助編寫原始程式碼文本、並編譯打包成為可用的程式，有些可以設計圖形介面。

▲Rstudio

機器學習應用於共用單車流量分佈

R 語言的一種的整合開發環境，可以輸出視覺化圖像。

▲開放數據(Open Data)

指的是一種經過挑選與許可的資料，這些資料不受著作權、專利權，以及其他管理機制所限制，可以開放給社會公眾，任何人都可以自由出版使用，不論是要拿來出版或是做其他的運用都不加以限制。

▲數據庫 (Database)

資料庫，簡而言之可視為電子化的檔櫃，存儲電子檔的處所，使用者可以對檔中的資料運行新增、截取、更新、刪除等操作。

▲數據挖掘 (Data mining)

是一個跨學科的電腦科學分支它是用人工智慧、機器學習、統計學和資料庫的交叉方法在相對較大型的資料集中發現模式的計算過程。資料採擷過程的總體目標是從一個資料集中提取資訊，並將其轉換成可理解的結構，以進一步使用。其目標是從大量資料中提取模式和知識，而不是挖掘資料本身。經常用於大規模資料或資訊處理（資料獲取、資料提取、資料存儲、資料分析和資料統計），還有決策支援系統方面的應用。

▲公共自行車 (Bike sharing system)

又稱共用單車，是一種能讓一般大眾共用自行車使用權的服務。可分為有樁和無樁兩種，前者取車和歸還的地點必須在樁所在處進行，後者在技術上則可以

機器學習應用於共用單車流量分佈

在任何可以定位到的地方進行取車和歸還的操作。

▲策略 (Strategy)

又稱博弈論 (game theory)，在賽局理論裡，玩家在賽局中的策略是指在所有可能發生情況下的一套完整行動計畫。玩家的策略會決定玩家在賽局的任一階段所採取的行動，不論這一階段之前是如何演變而來的。典型的策略、博弈論代表如囚徒困境 (Prisoner's Dilemma)。

第二章 文獻回顧

2.1 共用單車發展史

2.1.1 共用單車的起源

2.1.2 共用單車的發展

2.2 無樁共用單車使用流程

2.2.1 第一代無樁共用單車

2.2.2 第二代無樁共用單車

2.3 單車流量之統計與預測

2.3.1 有樁單車流量統計與預測之分區方式

2.3.2 無樁共用單車流量統計與預測之分區方式

2.4 聚類分析及其分支下三種分群法

2.4.1 K-平均演算分群法

2.4.2 密度分群法

2.4.3 DBSCAN 分群法

2.5 小結

第二章 文獻回顧

共用單車分為有樁、和無樁兩種模式。本文研究的對象為無樁模式的共用單車，無樁模式的共用單車具有取車和還車地點的隨機性，因此如何架構合理的單車存量分區方式、計算分區內的單車密度、並根據單車密度作為參數控制的條件、以提出相對應的單車疏散、與調度策略成為本研究的重點。

本章就共用單車發展史、無樁共用單車使用流程、以及單車流量之統計與預測三個方面的文獻與案例，收集本研究發展所需之條件。同時，於單車流量之統計與預測中，回顧聚類分析分支下的三種分類方法：K-平均演算分群法、密度分群法、以及 DBSCAN 分群法。

2.1 共用單車發展史

2.1.1 共用單車的起源

1965 年，荷蘭阿姆斯特丹一群異想天開的青年，將數十輛單車漆上白漆，放置街角供人們自行取用，他們稱此為「白色自行車計畫」(Witte Fietsen)。單車不上鎖，人們想騎就騎、想停哪就停哪。由此可見，最初的單車模式便是沒有固定停車區域和停車樁的設計，隨處可停，也就是今日共用單車採用的術語：「無樁式共用單車」。當年科技不成熟，計畫最終抵不過人性，單車多數不翼而飛，剩下的殘破不堪，計畫中途停止。

2.1.2 共用單車的發展

(1) 有樁共用單車出現和興起

過往因為技術問題，共用腳踏車的使用狀態無法被即時追蹤，為解決竊盜和破壞問題，建置停車樁、固定的停車區域再加上身分認證成為保障單車使用的方法，確保民眾會遵守共用規範的概念漸成主流，「有樁式共用單車」概念因此誕生。

1996 年，英國朴茨茅斯大學推出磁卡租借的共用單車系統，大幅改善遺失率；再隔兩年，法國雷恩市推出「智慧單車系統」(Smart Bike)，可在市內各處

機器學習應用於共用單車流量分佈

停放點取車還車，刷會員卡開鎖，每次最多使用三小時，一共兩百輛單車、二十五處停放點。

此後，歐洲數十個都市爭先恐後跨上單車，「有樁式共用單車」漸成流行。2007年，法國巴黎推出「自由單車系統」(Velib)，第一年有近三千萬名用戶，至2015年巴黎的共用單車超過兩萬輛，該系統也是現今發展最成熟的共用單車服務系統。臺灣 YouBike 也是採取此有樁式服務。

(2) 回歸無樁共用單車模式

中國各城市的有樁共用單車各有不同規範及收費標準，使用區域收到限制；使用者尋找固定停車點，仰賴各城鎮公共自行車站的網路佈局，常常找車比騎車所花費的時間來得要長；中國地廣，打樁、埋線設置的成本驚人，再加上後期管理不善，最終導致有樁共用單車使用率偏低。

為了真正解決城市裡面「最後一哩路」的交通問題，中國共用單車業者透過單車，把交通從固定節點的通勤，變成到任何地方都可以到達的便利交通網。無線網路的發達，物聯網、手機支付、定位技術、藍牙等新技術的成熟，讓沒有固定停放點、使用時間、隨取隨用的「無樁式共用單車」再也不會受到局限，無樁共用單車幾乎可以完美彌補中國現有共用單車遇到的所有問題。可以發現，無樁共用單車模式的出現是因為「最後一哩路」的需求和滯後發展的信用卡體系所導致的發達的移動支付浪潮兩個原因構成的。

2017 年 9 月，Mobike 進駐倫敦。2017 年 10 月，Mobike 進駐韓國水原市，華盛頓。2017 年 11 月，Mobike 進駐悉尼，鹿特丹。目前無樁共用單車進駐超過 20 座國際城市。

(3) 共用經濟模式產生的原因與發展

共用經濟最初是指透過網路技術刺激海量的閒置資源，讓物品的所有者通過某一平臺將物品的使用權暫時轉讓給他人，從而提高社會資源的利用率，為平臺兩端的參與者創造價值。在該模式中，企業透過打造平臺來連接閒置商品和人，而不是生產商品，我們可以將這種模式稱為 C2C 模式。其中，比較有代表性的企業有 Uber、Airbnb 等。而目前，共用經濟這一概念變得更加寬泛，除了 C2C 模式外還有 B2C 模式。這類公司與 C2C 模式公司的不同點在於其本身並不連接閒置資源，而是生產或購置大量商品投入市場，用於分時出租並收取租金，這本質上是租賃。較為典型的例子有共用單車、共用充電寶、共用雨傘、共用睡眠艙等。兩種模式各有優缺點，C2C 模式的優勢在於其體量價值，這類平臺希望連接的是整個社會的閒置資源，它所能觸及的體量遠超於 B2C 模式。而 B2C 模式的優勢在於其參與了從商品生產到線上線下運營的幾乎每一個環節，這可以保證所提供的服務品質標準化 (TechOrange, 2017)。

這種模式，意味著市場加速淘汰底層的公司，逼迫著經營體不斷地進行策略和服務上的改革，對於資料的分析顯得愈發重要，及時發現需求和問題，作出改善。

2.2 無樁共用單車使用流程

2.2.1 第一代無樁共用單車

使用方法：

獲得密碼：往 APP 輸入單車 id，獲得解鎖密碼

輸入密碼：轉動密碼鎖，輸入密碼

按鈕解鎖：機械按鈕解鎖

重新上鎖：轉到任意密碼，上鎖

操作 APP：按一下 APP 上的完成騎行按鈕

缺點：

需要改進的缺點：太繁瑣，解鎖和上鎖步驟過多；一旦密碼鎖被破壞，就失去功能，容易被偷。

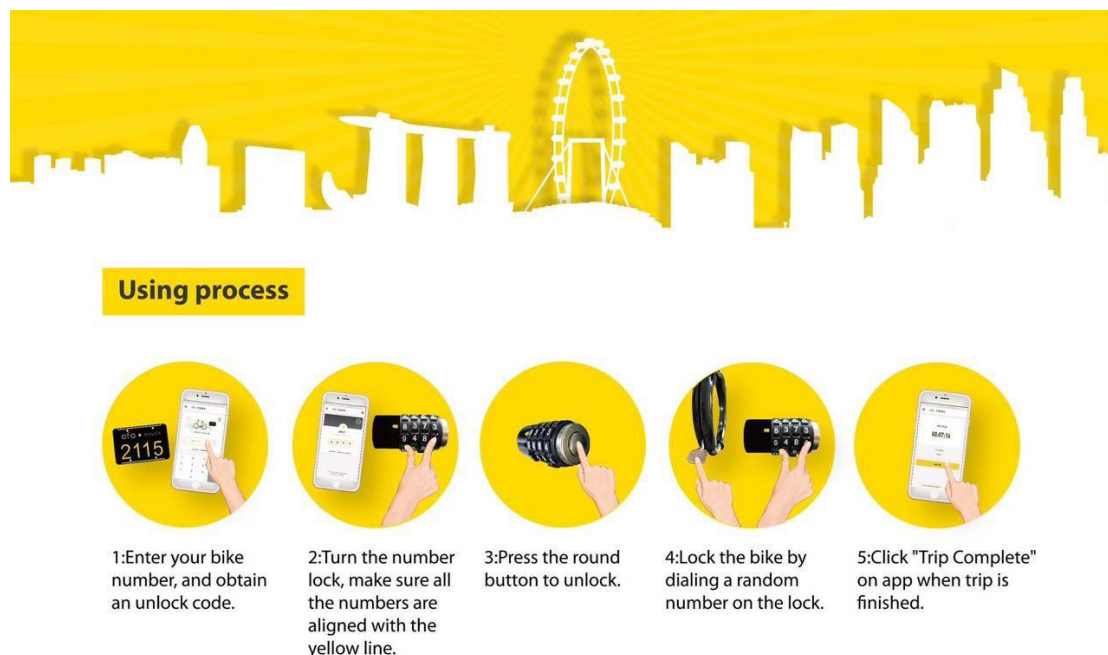


圖 2-1 第一代公用單車使用流程（來源：Ofo Official）

2.2.2 第二代無樁共用單車

使用方法：

解鎖：APP 掃取車身碼，自動解鎖，開始騎行。

上鎖：到達目的地後，將車輛停放在任何可合法停放的區域（非過道，馬路……），鎖車，APP 自動顯示騎行路線、長度、時間、費用。

使用原理：

拿摩拜舉例，由於新一代共用單車車身附有 GPS 定位系統，聯網模組，閒時會定時發送位置到公司雲端，使用時類似 SIM 卡實時發送位置到雲端；電池為

GPS、鎖車系統、聯網模組供電；而後輪的驅動可以為電池充電。

車輛防盜：

緊急情況下公司可以通過系統遠程鎖車，而對於破壞聯網功能的人，車輛本身又擁有一旦組件被開殼就向公司發送警報的設定，保證了車輛安全。

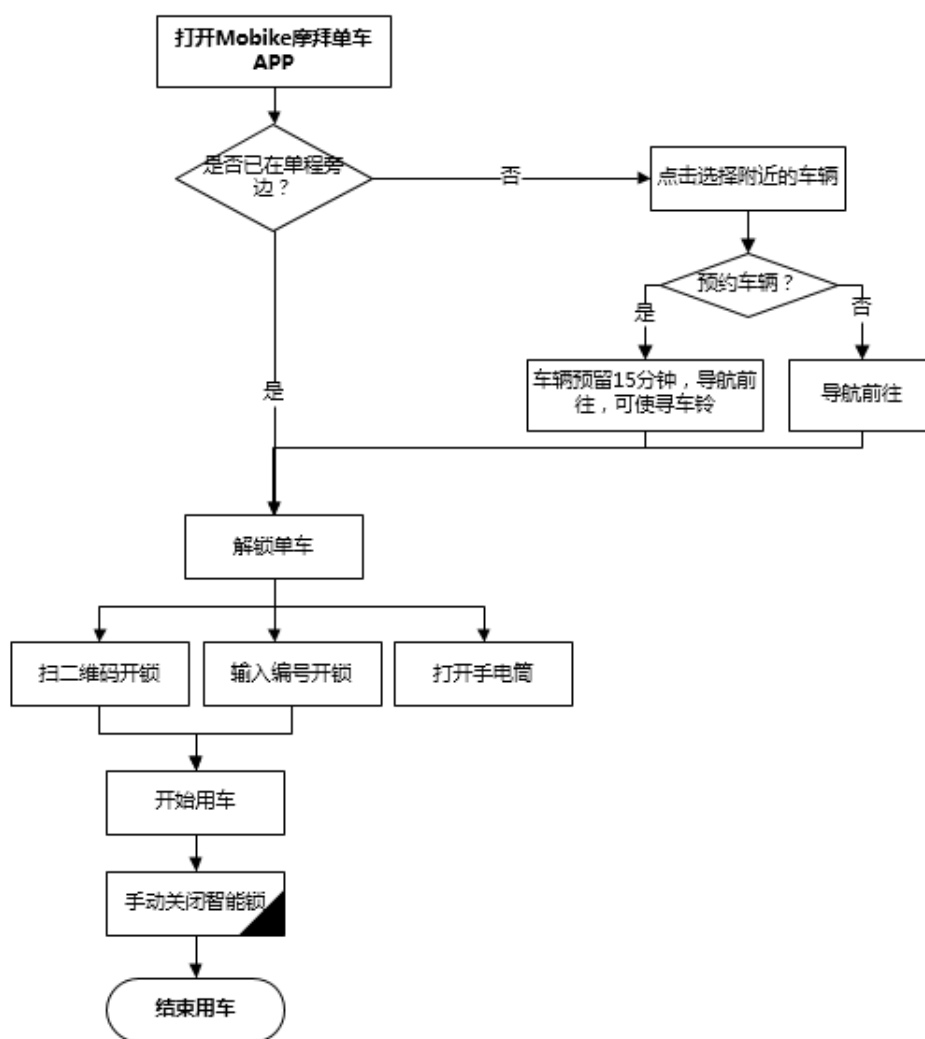


圖 2-2 第二代公用單車使用流程（來源：Mobike Global）

2.3 單車流量之統計與預測

本文研究的對象為無樁模式的共用單車，無樁模式的共用單車具有取車和還車地點的隨機性，因此如何架構合理的單車存量分區方式、計算分區內的單車密度、並根據單車密度作為參數控制的條件、以提出相對應的單車疏散、與調度策略成為本研究的重點。單車流量真實的統計和預測，建立在合理分區的基礎上，本節就有樁公用單車分區案例，整理單車流量之統計與預測需要面對的問題，以及應對的方法。整理如下：

2.3.1 有樁單車流量統計與預測之分區方式

要預測有樁模式的 共用單車在某一個具體時刻下，城市整體以及經過分群的樁點的每一個族群的單車數量，首先需要對單車樁點進行分群 (Li, 2015)。有樁模式的共用單車，單車的取得和還車地點被限定在樁點處，因此對固定樁位進行分群或者分區可以替代對單車進行的分群或者分區，計算同一群組中樁點處的單車取得和還車數量也較無樁模式的 共用單車方便和清楚。在一定區域範圍內，如果把 n 個樁分成 m 個群組 ($n > m$)，計算 m 個群組間單車的流動量，導入歷史資料，就可以分析未來的流動趨勢。如圖 2，呈現把樁分成 4 群後，模擬與預測群與群之間單車的流動機率。

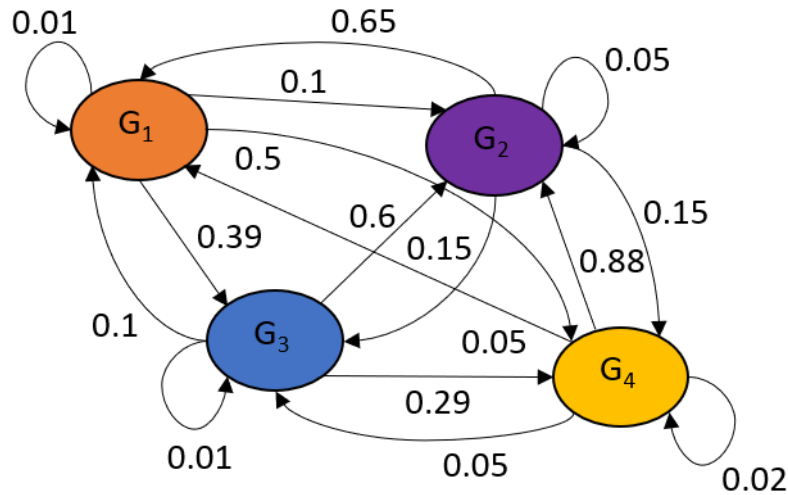


圖 2-3 群組間單車流動機率(改編自：Li, 2015)

同時，針對有樁的共用單車流量的統計與預測，包含分區方式，整理步驟，如下所示：

- (1) 通過訓練集求得一個預測模型，向這個模型輸入和測試集相同的特徵值，得到的結果和測試集的結果比較，得出誤差率 (Prediction error)。
- (2) 通過對比各種分區方式下的誤差率，選擇合適的分區方式以及參數設定。如用同一種分區方式進行分組，控制設定不同的參數，比較得出的誤差率，最終選擇合適的分區方式以及參數設定。

以下根據紐約市城市單車資料 (Citibike, 2017)，用不同的參數設定分區，得出的分區方式和結果：

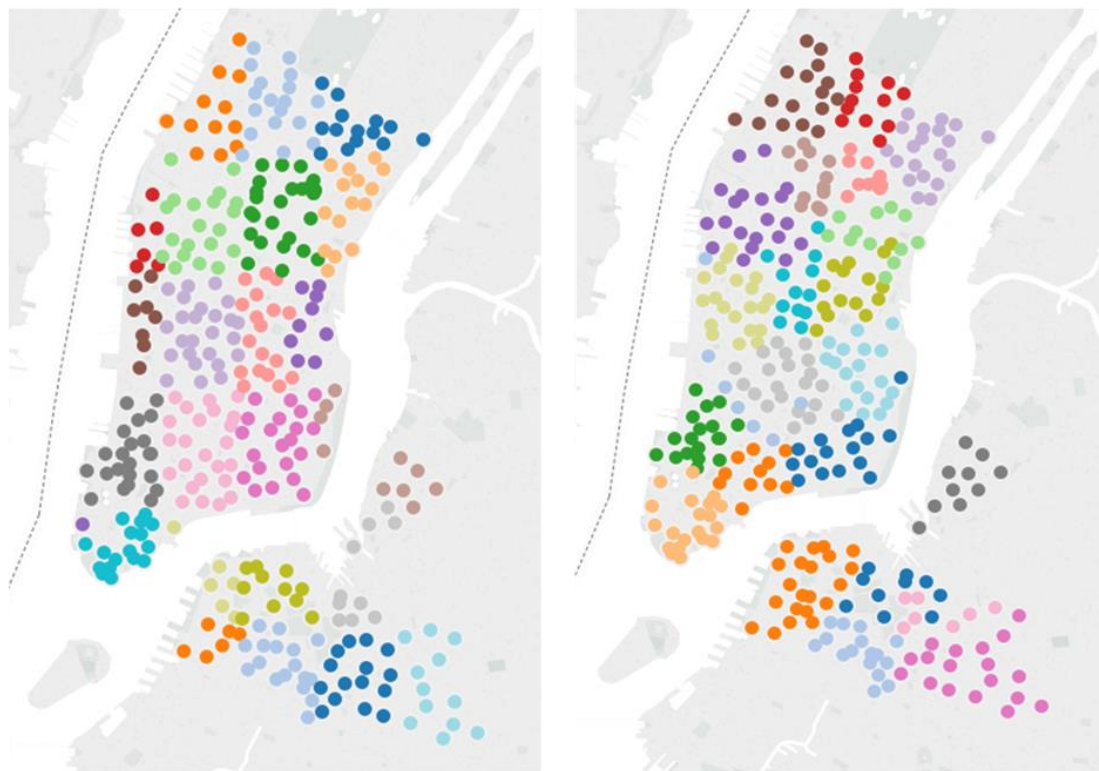


圖 2-4 紐約市城市單車分區示意圖 (來源: Li, 2015)

以及兩種不同設定的分區下，設定不同參數，最終得到的誤差率，如表 2-1、

表 2-2：

表 2-1 紐約市城市單車還車分區誤差率表 (來源: Li, 2015)

Check-out	All Hours		Anomalous Hours	
	GC	BC	GC	BC
HA	0.353	0.355	1.964	1.968
ARMA	0.346	0.346	2.276	2.273
GBRT	0.311	0.314	0.696	0.683

機器學習應用於共用單車流量分佈

HP-KNN	0.298	0.299	0.692	0.685
HP-MSI	0.288	0.282	0.637	0.503

表 2-2 紐約市城市單車取車分區誤差率表 (來源: Li, 2015)

Check-in	All Hours		Anomalous Hours	
	GC	BC	GC	BC
HA	0.347	0.352	1.837	1.835
ARMA	0.340	0.344	2.152	2.143
GBRT	0.309	0.309	0.681	0.671
HP-KNN	0.302	0.295	0.694	0.684
HP-MSI	0.297	0.290	0.642	0.506
P-TD	0.335	0.302	0.498	0.445

整理表格中所用到的所有分群方式，以及相關數據的解釋，結果如下：

HP：階級預測 (Hierarchical prediction)

BC：；流量矩陣階級分區 (bipartite clustering Transition-Matrix)

GC：地理分區 (Uniform geographical grid clustering)

機器學習應用於共用單車流量分佈

MSI：多重模擬基礎為影響 (multi-similarity-based inference) 建立的最佳模型

P-TD：最佳模型減去環境因數 (inter-cluster transition and trip duration)

HA：僅僅相似小時做參考，禮拜五的 1 到 2 點用所有 weekdays 的 1 到 2 點做參考

ARMA：禮拜五的 1 到 2 點用所有其他禮拜的禮拜五 1 到 2 點數據做參考

GBRT：僅僅由漸進梯度回歸樹模型 (GBRT) 計算，而不是整個城市使用量 (Entire Usage) * 比例 (proportion)

HP-KNN：階級預測 (Hierarchical prediction) - 最近鄰居法 (k-nearest neighbors algorithm)，一種和 HP-MSI 比較的演算方式

可以看到，此案例下預測系統的架構如圖 2-5 所示，由影響因素、資料結構、分區方式、與對應的策略這幾個部分組成：

機器學習應用於共用單車流量分佈

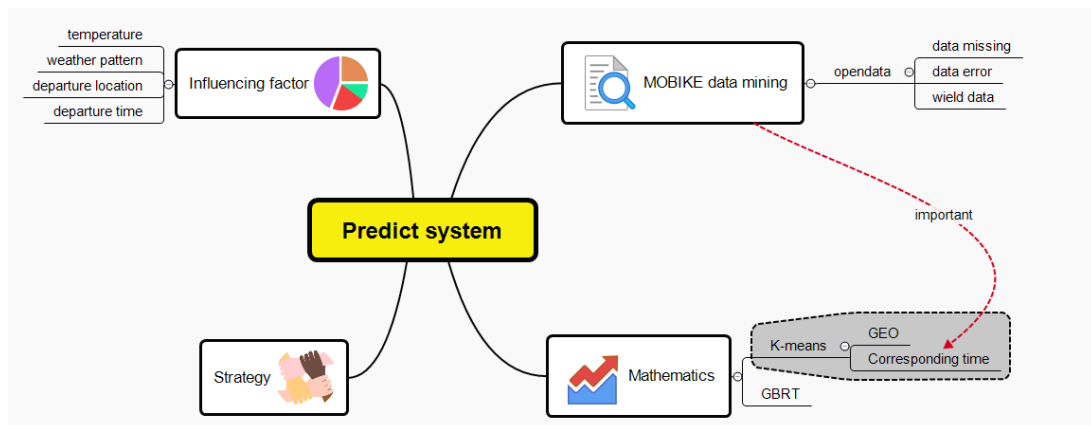


圖 2-5 紐約城市單車分區系統架構簡圖 (改編自：Li，2015)

2.3.2 無樁共用單車流量統計與預測之分區方式

根據對於有樁公用單車流量統計與預測之分區方式的探究，如若針對無樁公用單車，同樣用對樁進行分區的方式，則無法成立。但根據視對於區域的劃分為對樁的劃分，以劃分區域的方式進行分區，則不受樁的限制。而對於無樁公用單車，唯一可以定位的點位則是單車的取車點和還車點，對這兩個點的坐標進行劃分分區，同樣是一種可行的劃分分區的方式。

而對於點位的劃分，相對於樁的劃分，難度比較大。因為點的分佈屬於巨量分佈，同時具有不規則性，難以進行直接劃分。因此需要借用機器學習的方法來思考和分類。

2.4 聚類分析及其分支下三種分群法

聚類分析是機器學習無監督學習分支下的一種技術，聚類是把相似的物件通過靜態分類的方法分成不同的組別或者更多的子集 (Subset)，這樣讓在同一個子集中的成員物件都有相似的一些屬性。其中基於密度的聚類演算法，是為了挖掘有任意形狀特性的類別而發明的。此演算法把一個類別視為資料集中大於某閾值的一個區域。

以下整理兩種分區思維下所運用到的機器學習聚類分析原理、以及從屬聚類分析分支的分群方法。本研究使用到的聚類分析下的方法包括、但不限於：K-平均演算分群法，密度分群法、以及 DBSCAN 分群法。

2.4.1 K-平均演算分群法

K-平均演算分群法(k-means)，其過程包括：把 n 個點劃分到 k 個聚類中，使得每個點都屬於離他最近的均值（此即聚類中心）對應的聚類，以之作為聚類的標準，不斷重複運算，使每個點每次聚類時被劃分到它距離聚類中心最近的那個群。K-平均演算分群法對類球形且大小差別不大的類簇有很好的表現，但不能發現形狀任意和大小差別很大的類簇，且聚類結果易受雜訊資料影響。此外，K-平均演算分群法僅保證快速收斂到局部最優結果，從而導致聚類結果對初始代表點的選擇非常敏感（雷小鋒，2008）。對圖 4 左的點取 $K=3$ 的情況做分群，得到

圖 2-6 右的結果：

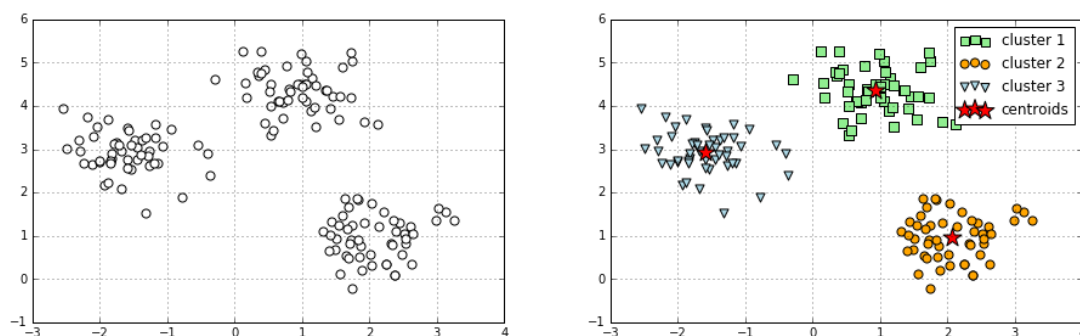


圖2-6 K-平均演算分群法演示

2.4.2 密度分群法

密度分群法是以密度為基礎 (Density-Based)：群集為一個物件密集的区域且由低密度的区域所圍繞(Richard, 2002)，這一類劃分不同密度集羣方法的統稱。本文所運用的密度分群法，應參考核密度估算 (Kernel Density Estimation)，該過程涉及根據離散事件(Discrete event)的已知位置來估計整個二維研究區域內的密度。核密度估算首先在研究區域頂部覆蓋一個網格，並基於每個網格單元的中心點計算密度估計。然後基於特定的核心函數(Kernel function)和帶寬(Bandwith)對離散事件與網格單元的中心點之間的每個距離進行加權。一般而言，核密度估算包含默認的帶寬演算法，這種計算預設半徑的方法通常會避免出現“點周圍的圓環”現象，該現象通常會隨稀疏資料集而產生。(Timothy, 2014)換句話說，密度分群法比較適用於劃分位於小尺度、且數據分佈較為稀疏的集羣。

同時，優化與歸納函數性質也是核密度估算的一大特性。它能把許多複雜資料簡化成近似的函數。如簡化直方圖，如圖 2-8 左。又如簡化離散數據點，如圖 2-8 右。

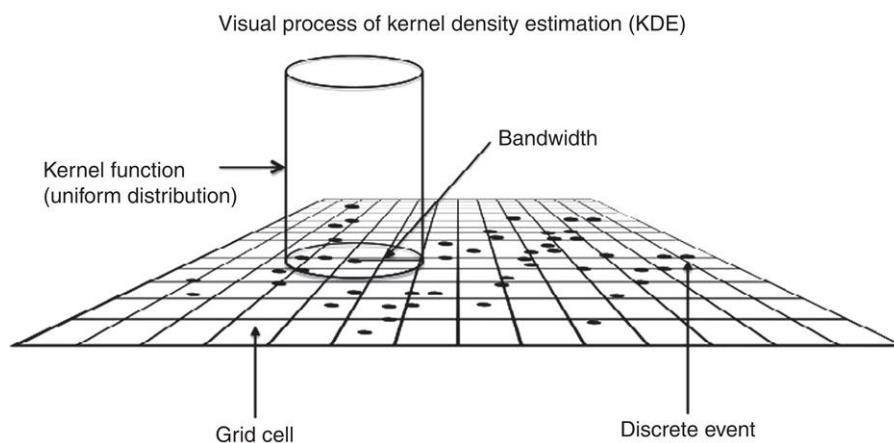


圖 2-7 核密度估算的視覺化 (Timothy, 2014)

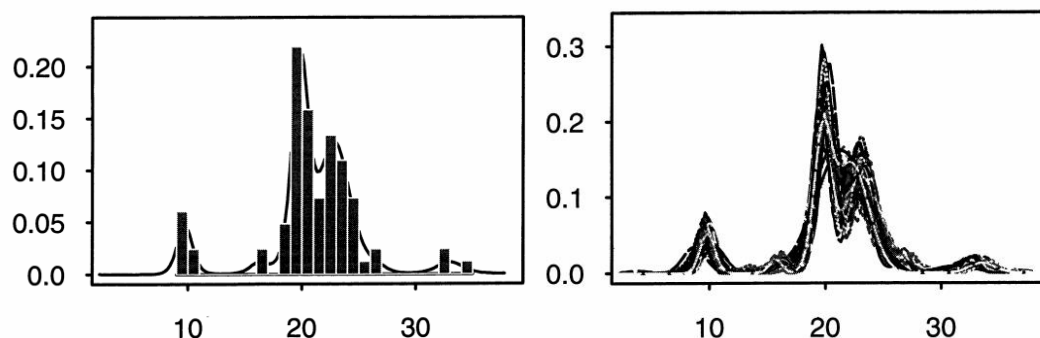


圖 2-8 核密度估算 (Michael, 1995)

2.4.3 DBSCAN 分群法

DBSCAN 分群法(Density-based spatial clustering of applications with noise)，這個演算法也是以密度為本的：給定某空間裡的一個點集合，這演算法能把附近的點分成一組，並標記出位於低密度區域的局外點。

DBSCAN 分群法之相關定義(Ester, 1996)： 1. 距離資料點半徑 ϵ 以內的鄰

近區域，則為該資料點的 ϵ -鄰近區域。 2. 資料點的 ϵ -鄰近區域中包含了至少最少個資料點(minPts)，則該資料點為核心物件。 3. 資料點 p 的位置是在某核心物件 q 的 ϵ -鄰近區域內，則資料點 p 被稱為由 q 直接密度可達(directly density-reachable)的物件。 4. 假如資料點 p 由 q1 直接密度可達、而 q1 由 q2 直接密度可達、……、而 qi-1 由 qi 直接密度可達，則資料點 p 被稱為由 qi 密度可達 (density-reachable)的物件，如圖 7 左，p 由 q 密度可達。 5. 假如資料點 p 和 q 都可由 o 密度可達，則 p 和 q 可以被稱為密度連接 (density-connected)。(圖 2-9 右)

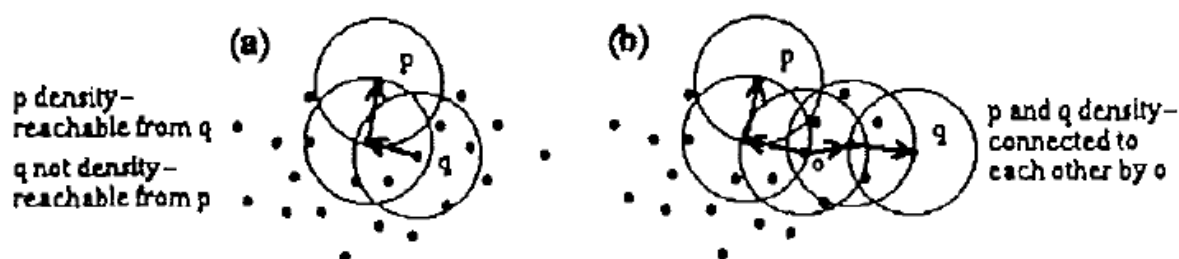


圖 2-9 DBSCAN 分群法原理 (Ester, 1996)

掃描半徑 (ϵ 比例值) 為圍繞中心點 (Core point) 的掃描半徑，最少點數 (minPts) 為使中心點不成為噪點 (Noise point, 即群外點) 的掃描區域內最少介面點 (Border point) 的個數。比如掃描範圍大於 3 個點時中心點才會被算作入群，不會被算作噪點。若同時調控掃描半徑和最少點數使得最終分群結果中的噪點最少，一般被認為是最佳的 DBSCAN 分群結果。

DBSCAN 分群法相對於其他種類以密度為基礎的分群方法，其優點在於：(一) 由於最少點數參數，單一連接現象 (single-link effect, 不同聚類以一點或極

少的線相連而被當成一個聚類) 能有效地被避免。(二) 能夠分辨噪點 (三) 只需掃描半徑與最少點數兩個參數, 對資料庫內的點的次序不敏感 (兩個聚類之間邊緣的點有機會受次序的影響被分到不同的聚類, 另外聚類的次序會受點的次序的影響)。

2.5 小結

藉由文獻回顧，可瞭解對於有樁公用單車流量統計與預測之分區方式，對樁分區的方式是最為合理的，因為樁有坐標確定的優點。與此同時，單車的取車和還車都被限定在樁點，意味著假若忽略正在被使用的單車，對樁的分類可視作對單車的分類。

而針對無樁公用單車，若同樣用對樁進行分區的方式，則無法成立。借鑒有樁公用單車視對樁的劃分，總結出兩種劃分方法。第一種方法，視對於區域的劃分為對樁的劃分，以劃分區域的方式進行分區，不受是否曾在樁的限制，如圖 2-10 所示。第二種方法，對於無樁公用單車，唯一可以定位的點位則是單車的取車點和還車點，對這兩個點的坐標進行劃分分區，同樣是一種可行的劃分分區的方式。

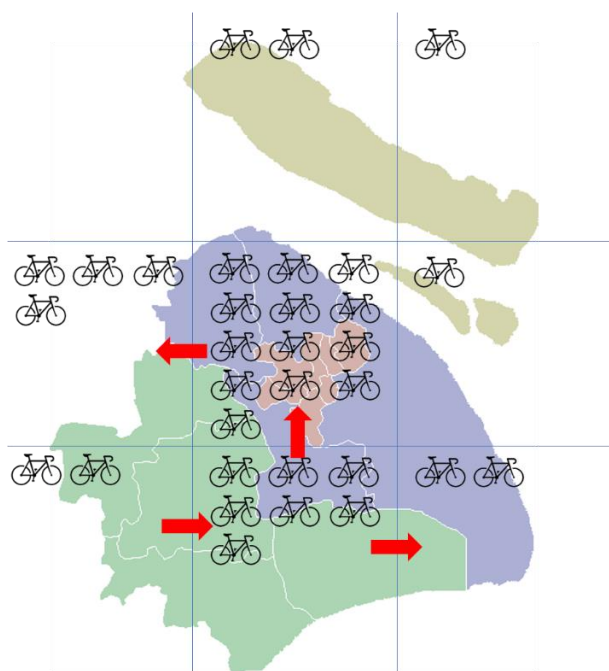


圖 2-10 區域劃分分區示意圖

對於以上兩種方法可以用到的聚類分析分支下的三種分類方法：K-平均演算分群法、密度分群法、以及 DBSCAN 分群法，各自有合適的實際使用情境。K-平均演算分群法對類球形且大小差別不大的類簇有很好的表現，但不能發現形狀任意和大小差別很大的類簇。密度分群法比較適用於劃分位於小尺度、且數據分佈較為稀疏的集群。DBSCAN 分群法相對於其他種類以密度為基礎的分群方法，則可以避免單一連接現象、以及分辨噪點，對於處理大尺度下的數據分佈沒有明顯弱點。

第三章 理論與方法

3.1 基礎工具和程式語言

3.2 基於閾值控制思維的逆向流動獎勵制度及其沙盤推演

3.2.1 基於閾值控制思維的逆向流動獎勵制度

3.2.2 獎勵機制的沙盤推演

3.3 小結

第三章 理論與方法

單車業者具備及時、或者是超前調度的管理能力十分重要。本章節就模擬化的獎勵機制及其沙盤推演為基礎，闡述實際運用獎勵機制時要面對的邏輯問題，及對應的策略與方案。以下說明基礎工具和程式語言、基於閾值控制思維的逆向流動獎勵制度及其沙盤推演：

3.1 基礎工具和程式語言

本研究以 R 程式語言作為處理和分析數據的工具。R 語言，是一種以命令列指令為主，呼叫函數處理數據、統計分析、輸出圖形的軟體程式語言。近年來大數據 (Big Data) 的興起，使 R 語言在資料學界備受重視。RStudio 則是編寫 R 語言程式的一種整合開發環境，它是開源的自由軟體，具備許多編碼工具與專案式管理介面，方便整理、迅速瀏覽檔案與功能集。除此之外，它可將 R 語言編寫的程式輸出為 HTML、PDF、Word 檔、投影片，支援 ggvis 等包裹(package)以產生可視化的互動圖型。(Lander, 2013)，如圖 3-1 所示。

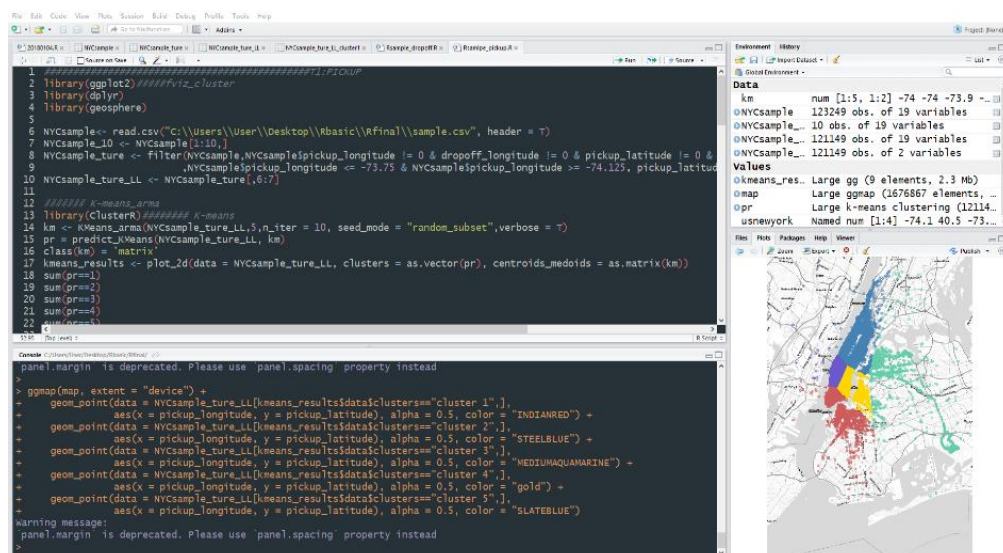


圖 3-1 Rstudio 的使用者介面

3.2 基於閾值控制思維的逆向流動獎勵制度及其沙盤推演

本研究基於區塊內單車在單位時段內的取得量和還車量，預測相鄰時段內單車騎行目的地區塊的趨勢分佈，並提出基於閾值控制思維的逆向流動獎勵制度作為單車的疏散與調度策略。本文希望通過獎勵騎行者的方式來控制每個分區的單車存量。以下就基於閾值控制思維的逆向流動獎勵制度、及其沙盤推演兩個方面，整理並解析出相關的理論與方法：

3.2.1 基於閾值控制思維的逆向流動獎勵制度

本文希望通過獎勵騎行者的方式來控制每個分區的單車存量。理論上應得知每個分區內單車的 GPS 資訊，取得單車存量數量後進行分析。此種資料大多為商業機密資料，本研究採用少數對外開放的上海開放資料創新應用大賽（Soda Opendata, 2018）於 2016.8 發佈的 101843 筆騎行資料。其中包含訂單號、單

機器學習應用於共用單車流量分佈

車編號、使用者編號、取車時間、取車經度、取車緯度、還車時間、還車經度、還車緯度以及軌跡。圖 3-2 共用單車控制系統模型，系統將單位時間內、單位區域取車 (Check-in) 的次數視為該區域人對車需求；而單位時間、單位區域還車 (Check-out) 次數視為對該區域車供給。為了保證每一個區塊內單車存量處於良好的狀態(圖中介於上下觸發獎勵機制線間，即為正常區間)，通過獎勵或強制辦法控制取車與還車的速度。當存量過多、且高於最高閾值；或者當存量過少、且低於最低閾值的時，系統將觸發獎勵機制，給予紅利或者折扣，以獎勵下一個時段內還車、或者取車的車輛，從而使存量結果較未實行獎勵機制前相比更趨向於良好狀態。然而，當系統顯示單車密度供給過量或者缺乏時，則只好調派卡車強制運輸，強制使系統再度回歸到正常範圍。

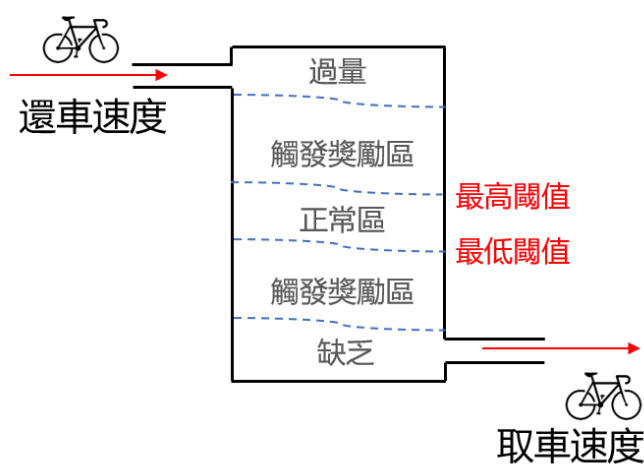


圖 3-2 共用單車控制系統模型

3.2.2 獎勵機制的沙盤推演

根據上述單車控制系統模型，本節進行理論沙盤推演。考量到對局域進行劃

機器學習應用於共用單車流量分佈

分的情形下，要使每一塊局域都被劃入對應的組別，矩形的劃分方法最簡單也最為實用。假設格子分區為最合適的分區方式，街道地圖被分成 9 個方正的矩形(圖 3-3)。

對於格子的尺寸，考量到無樁模式的共用單車時，在進行分群之前，還需考慮使用者在分區範圍內，是在步行的舒適條件下，容易找到可使用的 共用單車為前提。然而最適的步行距離、受到區域的氣候條件、垂直距離，以及城市經濟發展狀況的影響 (Speck, 2012)。對於上海市社區居民的考量，在 787m 的步行距離以內，居民選擇步行出行的可能性超過 10%，而在 787m 之後，選擇步行的比例很低。在進行步行距離的調查時，沒有考慮往返距離對結果的影響，這會使步行距離的閾值估計值偏大 (王寧, 2015)。由於本研究所採集的地理訊息資料相關資料，其精確度均取值到百位數，為簡化計算，本文取城市最適步行距離範圍的整數值：700m 作為後文分區演練的基準值，它約略為 10 分鐘步行距離，可代表局部日常步行活動情況。除此之外，以下均以單車表述無樁的共用單車。

在一個分區內，使用者在尋找到可用單車前，步行距離應以不超過最適步行距離為限，即取等邊矩形對角線的最小值為 700m，對應的邊長為 550m。下文統一用 550m 作為分區邊長進行演練。

本節用數字表示分區內單車的密度。假設一開始(2016.8.1 的 00:00) 為初始平衡狀態，單位區域內的單車密度皆為 6 單位元元密度(圖 3-4 左)，經過一小時後(01:00 時刻)，單車存量的單位元元密度產生了數值變化(圖 3-4 右)。其

機器學習應用於共用單車流量分佈

中每個單位區域內，單車還車速度 (check out) 為圖 3-5 左、以及單車取車速度 (check in) 為圖 3-5 右。

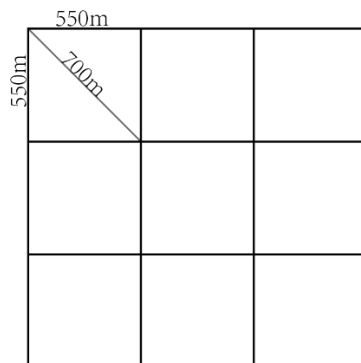


圖 3-3 沙盤分區示意圖

2016.8.1 的 00:00, 01:00 狀態分別為：

6	6	6	6	<u>5</u>	6
6	6	6	<u>5</u>	10	<u>5</u>
6	6	6	6	<u>5</u>	6

圖 3-4 分區的單車單位元元密度

假設 00:00 的狀態為初始平衡狀態，01:00 時刻時我們通過過去一個小時內

還車

1	<u>1</u>	2
<u>2</u>	6	<u>1</u>
2	<u>1</u>	4

取車

1	<u>2</u>	2
<u>3</u>	2	<u>2</u>
2	<u>2</u>	4

取車和還車的記錄得到如下數據，如圖 3-5 左為單位時間內還車量，圖 3-5 右為單位時間內取車量（劃線部分取車量大於還車量）：

圖 3-5 分區的單車還車速度 (check out)(左)與取車速度 (check in)(右)

根據上述，設計取車的獎勵公式(1)、和還車的獎勵公式(2)（其中 X 為取車速度，Y 為還車速度，t 表單為時段）：

假設單車單位元密度過多、且高於最高閾值時；

且 $X_{t-1} < Y_{t-1}$ 時，則取 $|X_t - Y_t|/2^{X_t}$ 作為取車(check in)的獎勵額度 (1)

根據獎勵公式(1)，系統為了促使狀態向初始時刻靠攏，會獎勵”取車”的每一輛車的獎勵數值為： $|取車量-還車量|/2^x$ ，換句話說這個時段內的第一輛取車得到的獎勵是最大的，隨後的每一輛單車的獎勵會隨著取車的單車數量的增加而減少，每取車一輛，則獎勵為 $|取車量-還車量|$ 除以(2的增量數方根)而遞減： $1/2$ 、 $1/2^2$ 、 $1/2^3$ 、以此類推，最後取車的獎勵無限趨向為0，但始終不為0。

假設單車單位元密度過少、且低於最低閾值時；

且 $X_{t-1} \geq Y_{t-1}$ 時，則、取 $|Y_t - X_t|/2^{Y_t}$ 作為還車(check out)的獎勵額度 (2)

根據獎勵公式(2)，系統為了促使狀態向初始時刻靠攏，會獎勵 t 時段內還車的每一輛單車的獎勵數值為： $|還車量-取車量|/2^y$ ，換句話說這個時段內的第一輛還車的單車得到的獎勵是最大的，隨後的每一輛單車的獎勵隨著”還車”的單車數量的增加而減少，每”還車”一輛，則獎勵為 $|還車量-取車量|$ 除以(2的增量數方根)而遞減： $1/2$ 、 $1/2^2$ 、 $1/2^3$ 、以此類推，最後”還車”的獎勵無限趨向為0，但始終不為0。

例如從01:00到02:00每輛單車取車和還車得到的獎勵為圖12左，對應的獎勵公式為圖12右（其中 X 為取值，Y 為還值，劃線四格實行還車獎勵，剩餘五格實行取車獎勵）：

機器學習應用於共用單車流量分佈

$(+1-1)/2^x$	$(+1-2)/2^y$	$(+2-2)/2^x$	$(+X_{t-1}-Y_{t-1})/2^{xt}$	$(+X_{t-1}-Y_{t-1})/2^{yt}$	$(+X_{t-1}-Y_{t-1})/2^{xt}$
$(+2-3)/2^y$	$(+6-2)/2^x$	$(+1-2)/2^y$	$(+X_{t-1}-Y_{t-1})/2^{yt}$	$(+X_{t-1}-Y_{t-1})/2^{xt}$	$(+X_{t-1}-Y_{t-1})/2^{yt}$
$(+2-2)/2^x$	$(+1-2)/2^y$	$(+4-4)/2^x$	$(+X_{t-1}-Y_{t-1})/2^{xt}$	$(+X_{t-1}-Y_{t-1})/2^{yt}$	$(+X_{t-1}-Y_{t-1})/2^{xt}$

圖 3-6 獎勵機制示意

以圖 3-4、圖 3-5 位於正中的格子紅色標識的資料為例，00:00 時單車密度為 6，01:00 時單車密度為 10，根據 00:00 到 01:00 間的單車流動資料，單車還車 6，取車 2，變化的過程為：

$$6 (\text{初始}) + 6 (\text{還車}) - 2 (\text{取車}) = 10 (\text{最終})$$

因為 $2(\text{取車}) < 6(\text{還車})$ ，即符合 $X_{t-1} < Y_{t-1}$ 時，採取車獎勵公式(1)，所以在 01:00 到 02:00 間每輛單車，還車於此區域的取車的取車獎勵為： $|2-6|/2Y$ ，Y 為取車量的順序編號。

第一輛取車獎勵額： $|2-6|/21 = 2$ ，

第二輛取車獎勵額： $|2-6|/22 = 1$ ，

第三輛取車獎勵額： $|2-6|/23 = 0.5$ ，

第四輛取車獎勵額： $|2-6|/24 = 0.25$

.....

因代替了原本需要用卡車運輸調度產生的費用，獎勵額度可以看作這些費用的百分比。比如原本這一區塊需要通過卡車運輸調度的費用為 F 元，即平均每小

機器學習應用於共用單車流量分佈

時 $F/24$ 元，從 00:00 到 01:00 總共產生的獎勵額度為每個方格內單車變化量，

即把圖 8 右格子中的數值，減去圖 8 左對應的格子中的數值，計算每一項的和，

如圖 3-7：

$ 6-6 =0$	$ 5-6 =1$	$ 6-6 =0$
$ 5-6 =1$	$ 10-6 =4$	$ 5-6 =1$
$ 6-6 =0$	$ 5-6 =1$	$ 6-6 =0$

圖 3-7 總獎勵額度計算

總共為 $1+1+1+1+4=8$ ，所以上文中間格子相同編號的單車獲得的實際獎勵為：

$$\text{第一輛取車獎勵： } 2/8 \times 4 \times F/24 = F/24 \text{ 元}$$

$$\text{第二輛取車獎勵： } 1/8 \times 4 \times F/24 = F/48 \text{ 元}$$

$$\text{第三輛取車獎勵： } 0.5/8 \times 4 \times F/24 = F/96 \text{ 元}$$

$$\text{第四輛取車獎勵： } 0.25/8 \times 4 \times F/24 = F/192 \text{ 元}$$

..... 以此類推。

3.3 小結

在基於閾值控制思維的逆向流動獎勵制度及其沙盤推演的演示下，本章節就獎勵機制策略及其合理性做探討，企圖強化從策略到實際運用之間的邏輯性，提供參與者模擬化獎勵機制的體驗機會，建立初步的獎勵機制模型及其運作原理。

實際運用獎勵機制時要面對的邏輯問題，包含基建立於逆向疏散獎勵的邏輯性、如何最大化激發使用者的調度興趣、以及合理的獎勵額與獎勵元數的分配。以下就這三個方面做小結：（一）以逆向疏散獎勵的邏輯性為基礎，系統將單位時間內、單位區域取車（Check-in）的次數視為該區域人對車需求；而單位時間、單位區域還車（Check-out）次數視為對該區域車供給。當存量過多、且高於最高閾值；或者當存量過少、且低於最低閾值的時，系統將觸發獎勵機制，給予紅利或者折扣，以獎勵下一個時段內還車、或者取車的車輛，從而使存量結果較未實行獎勵機制前相比更趨向於良好狀態。（二）根據激發使用者的調度興趣，建立隨車輛順序變化，由多到少的獎勵額度。即區域內單車存量距離平衡狀態的差值，與獎勵額成正比。（三）根據合理獎勵額與獎勵元數的分配，建議將原本人力運輸的成本作為獎勵的總元數，按每輛車的獎勵額所佔總獎勵額的百分比，分配給每輛單車。

第四章 實作分析

4.1 階段一：調取並處理數據

4.2 階段二：探究變量對於預測系統的影響

4.2.1 計程車上客點位分析

4.2.2 單車數據資料精確度

4.3 階段三：分群方法的探究及其實際運用

4.3.1 應用密度分群法，處理較小的街道尺度，以上海靜安寺及其周邊為例

4.3.2 應用 DBSCAN 分群法，處理較大的城市尺度，以上海四環及其內部為例

4.4 階段四：疏散，調度策略總結

4.5 小結

第四章 實作分析

本研究主張應用機器學習聚類分析於單車流量預測與調度。在應用程式的開發上，本研究採用 R 語言、並使用” RStudio” 軟體作為編寫 R 語言程式的整合開發環境。而機器學習的聚類分析 (cluster analysis) 則採行三種方法以進行測試、與比較分析。三種方法分別是：K-平均演算分群法 (K-means)、密度分群法、以及 DBSCAN 分群法 (Density-based spatial clustering of applications with noise)。

本研究分別以上海市靜安寺周邊、以及上海四環 (121.75°E-121.125°E, 30.95°N-31.45°N) 為兩種不同空間尺度的研究範圍，以探討在不同空間尺度下，記錄摩拜單車 (Mobike, 2018) 在單位時段內的取得量和還車量，借鑒紐約市計程車在聚類分析下的分區方式，以找出最佳的上海單車的的分群法、與分區情境的對應模式。在最佳分群與分區對應的模式下，以預測相鄰時段內單車騎行目的地區域的趨勢分佈為基礎，提出基於閾值控制的逆向流動獎勵制度，作為單車的疏散與調度策略。

根據上述理論與方法，本研究實作分為：調取並處理需要的數據、探究變量對於預測系統的影響、探究分群的方法、建立疏散和調度策略、等四個階段進行，如下說明：

4.1 階段一：調取並處理數據

其任務為尋找公開的 共用單車流量開放數據，以及用 R 處理，錯誤的、丟失的、超過範圍的數據。截至 2016.8，上海市預計投放使用的摩拜單車使用數量達到 223,000 輛左右(Soda Open data, 2018)，中國每一輛摩拜單車平均每天被 5.4 個人騎行（北京清華同衡規劃設計研究院，2017），可以粗略得出 2016.8 上海市的摩拜單車每個月有約 36,000,000 筆出行記錄。根據上海市城區的經緯度位置以及每個經緯度間隔 111,000m 的常識，加上城市最適步行距離範圍值為 700m，即格子對角線長為 550m，估計得出經緯度 550m*550m 的格子範圍內單車的平均分佈量為約 2857 輛（不減去淘汰、折損、返修的車）。若在市率 50%計算，則約為 1428 輛。以下實作部分均採用每個 550m*550m 方格內單車平均存量為 1428 輛作為 2016.8 上海市單車分佈的情形。

對比上海開放資料創新應用大賽 2016.8 的 101843 筆資料，可以知曉所抽樣的資料約占當月騎行量的 0.28%，在作資料處理的時候採用倍數放大法按比例放大資料量。

根據得到的數據，本節運用 R 程式語言，通過 Rstudio 軟體支持，建立視覺化腳本，運用視覺化資料的技術，來展現數據處理後的實際分佈情況。

如下圖，在禮拜一工作日的過程中，00:00 至 06:00 借出非常少。06:00 至

機器學習應用於共用單車流量分佈

12:00 借出，12:00 至 18:00 借出，18:00 至 24:00 借出分佈漸漸增多且覆蓋範圍

圍有些許的增加，範圍上沒有明顯的變化。

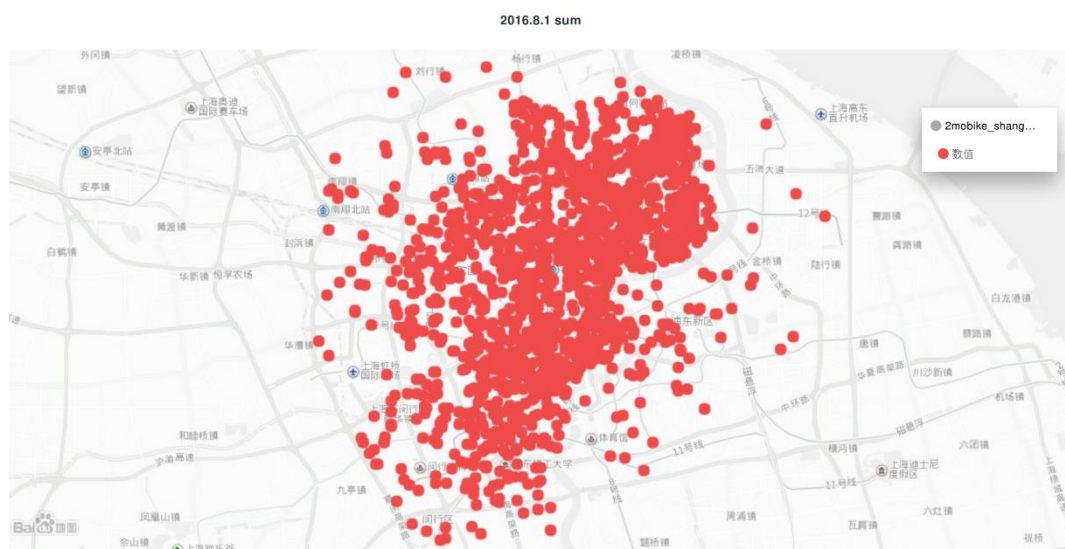


圖 4-1 上海市 2016 年 8 月 1 日（禮拜一）00:00 至 24:00 取車點示意圖



圖 4-2 上海市 2016 年 8 月 1 日（禮拜一）00:00 至 06:00 取車點示意圖



圖 4-3 上海市 2016 年 8 月 1 日 (禮拜一) 06:00 至 12:00 取車點示意圖



圖 4-4 上海市 2016 年 8 月 1 日 (禮拜一) 12:00 至 18:00 取車點示意圖



圖 4-5 上海市 2016 年 8 月 1 日（禮拜一）18:00 至 24:00 取車點示意圖

如下圖，在禮拜四、五，同樣時間段，於週末的關係，分佈點的數量和範圍明顯變大了。

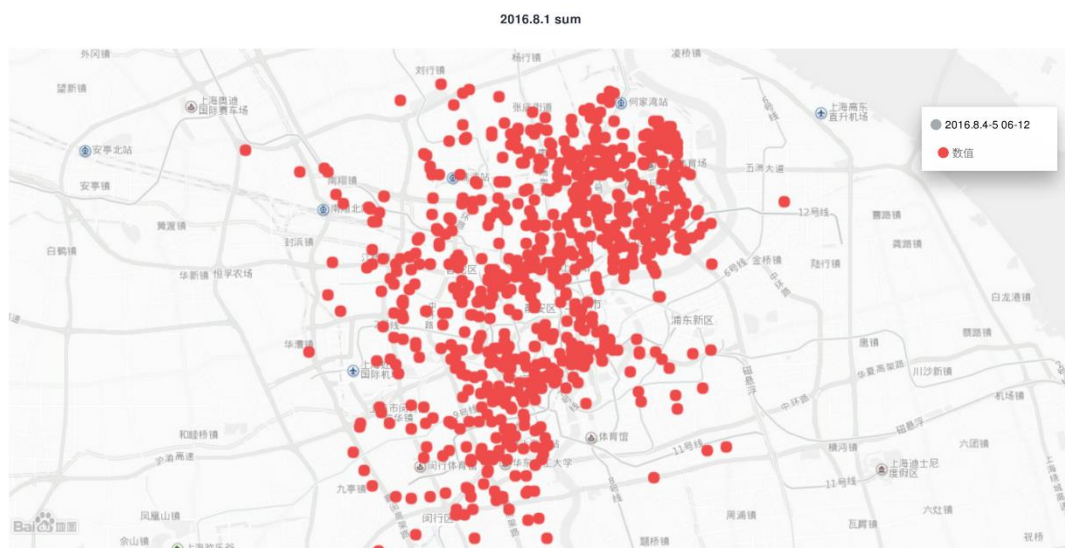


圖 4-6 上海市 2016 年 8 月 4.5 日（禮拜四、五）取車點示意圖

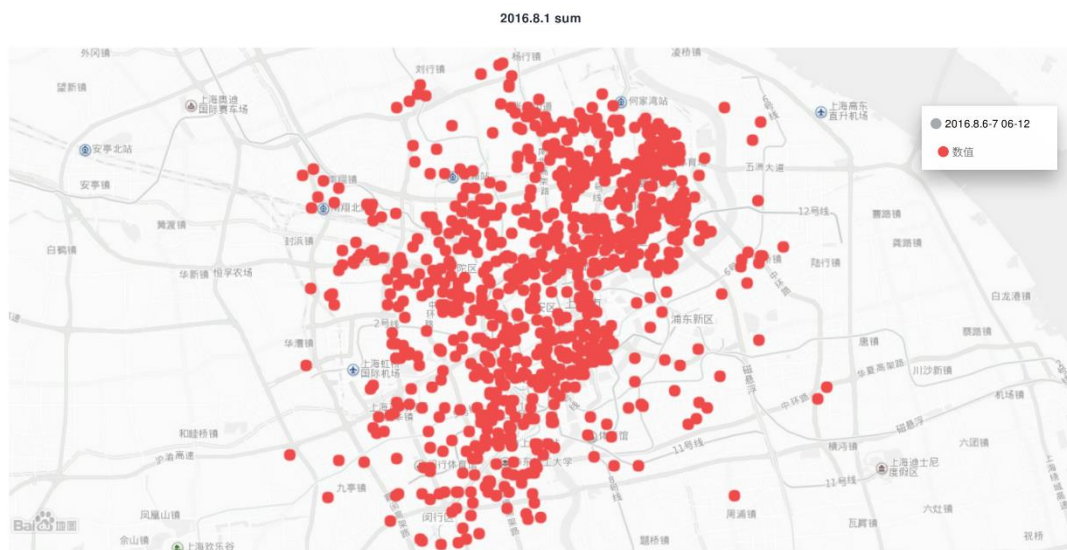


圖 4-7 上海市 2016 年 8 月 6.7 日（禮拜六、日）06:00 至 12:00 取車點示意圖

根據下圖對於點位的熱力分析結果，可以更明顯察覺到範圍的擴大。



圖 4-8 上海市 2016 年 8 月 4.5 日（禮拜四、五）00:06 至 00:12 取車點熱力示意圖

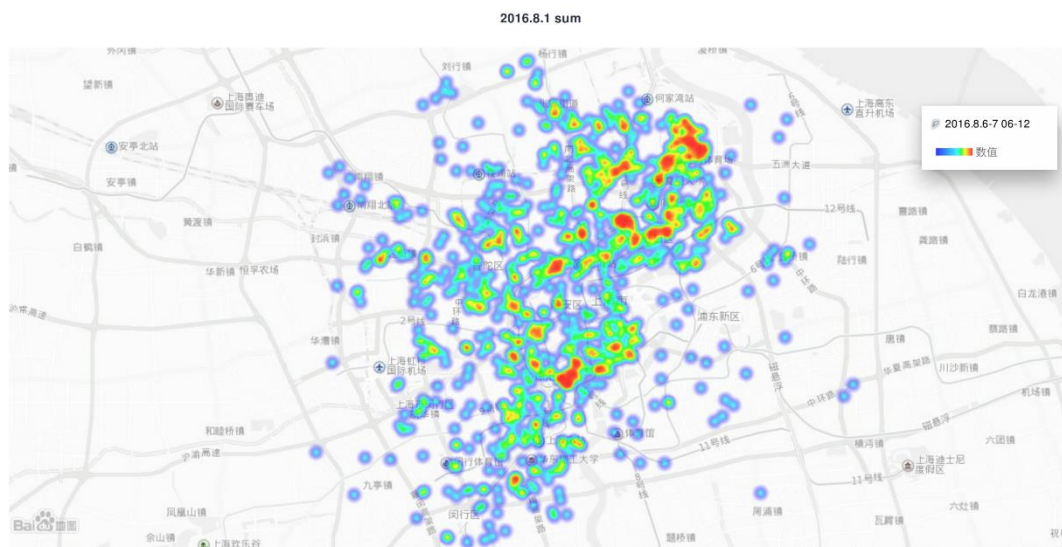


圖 4-9 上海市 2016 年 8 月 6.7 日 (禮拜六、日) 00:06 至 00:12 取車點熱力示意圖

4.2 階段二：探究變量對於預測系統的影響

其任務為考量數據精確度對於研究結果的影響。本節通過對紐約市計程車上客點的視覺化分析，考量對於位於不同尺度下的數據資料，應採取的聚類分析的方法。

4.2.1 計程車上客點位分析

針對紐約市計程車2015.7.23的上客地點做統計與視覺化，對於計程車，點位的分佈可以跨越很大的距離，因為計程車在地圖上的移動距離不向單車一樣受到人力的限制。

4.2.2 單車數據資料精確度

經過對上海開放資料創新應用大賽(Soda Open data, 2018) 2016年8月的101843筆騎行資料的分析，得出圖示結果：

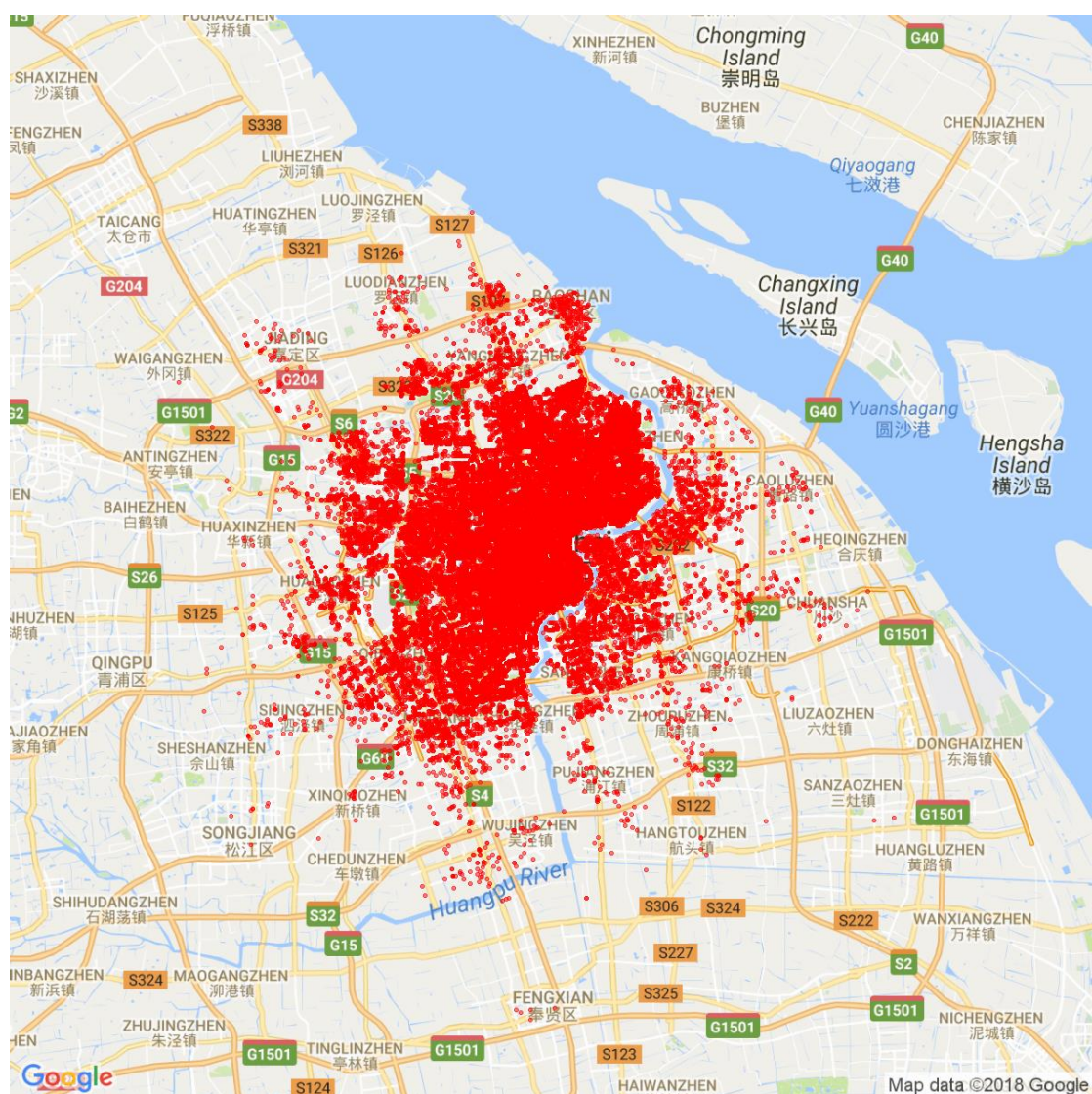


圖 4-12 上海摩拜單車還車點分佈圖 a

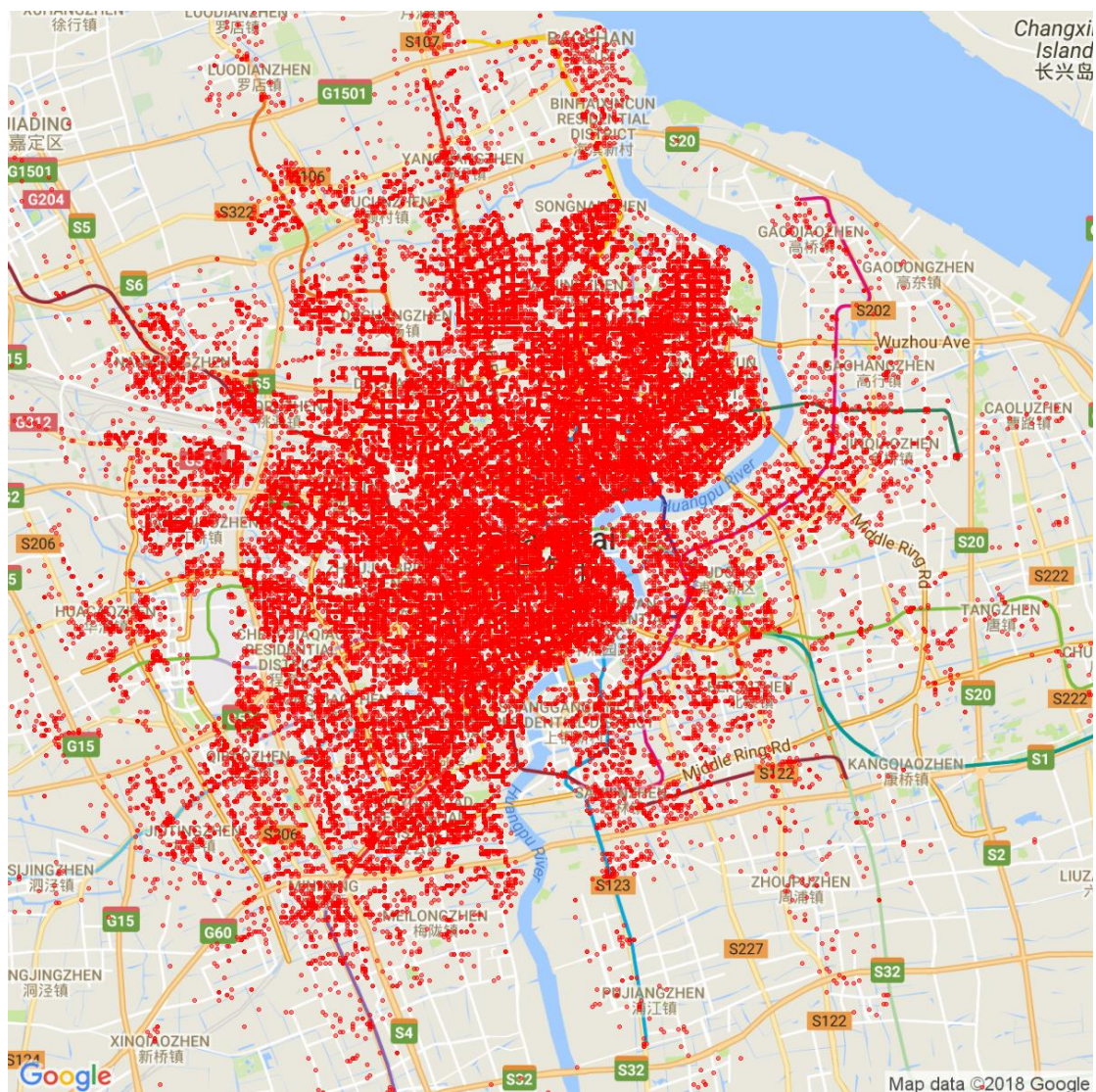


圖 4-13 上海摩拜單車還車點分佈圖 b

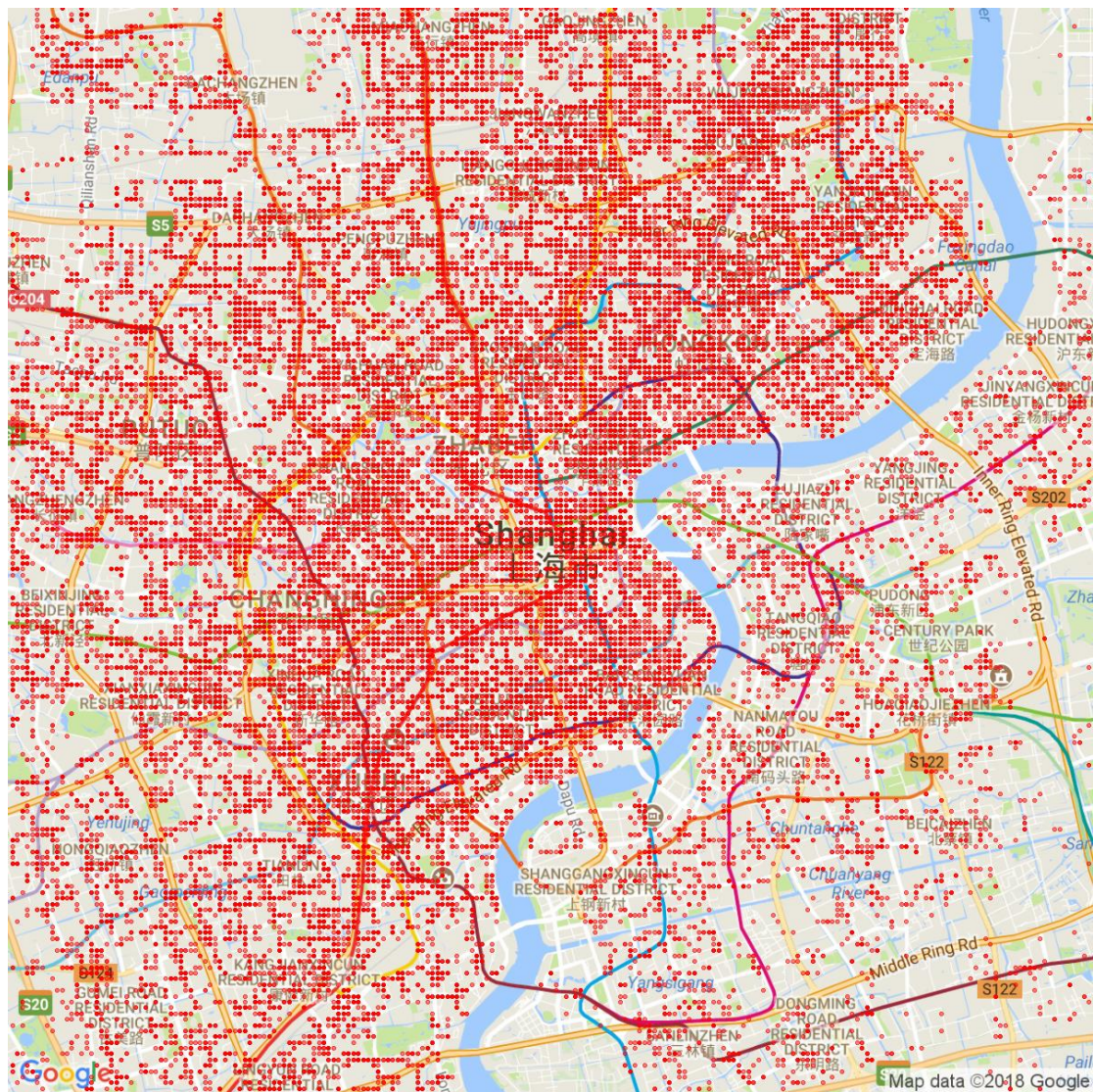


圖 4-14 上海摩拜單車還車點分佈圖 c

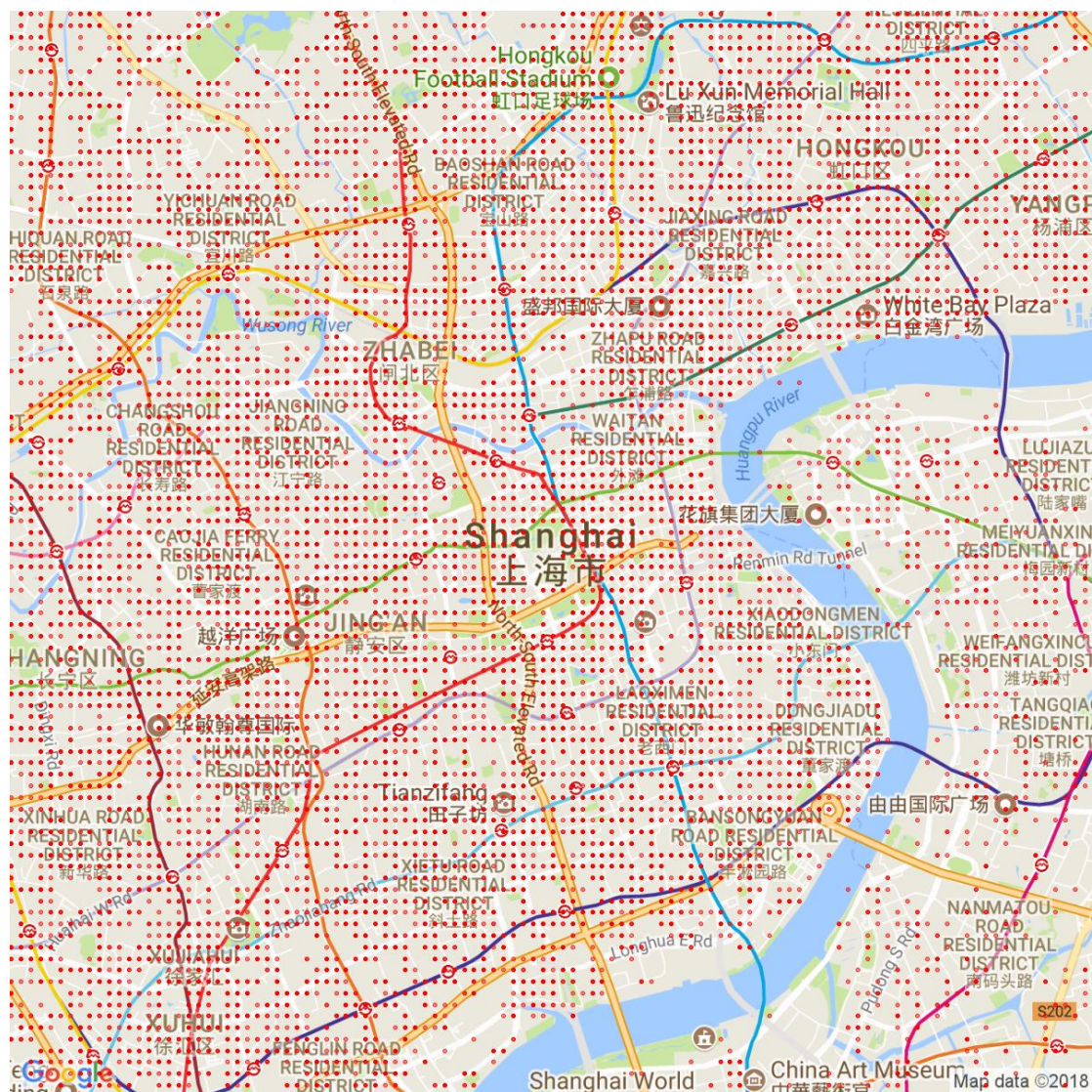


圖 4-15 上海摩拜單車還車點分佈圖 d

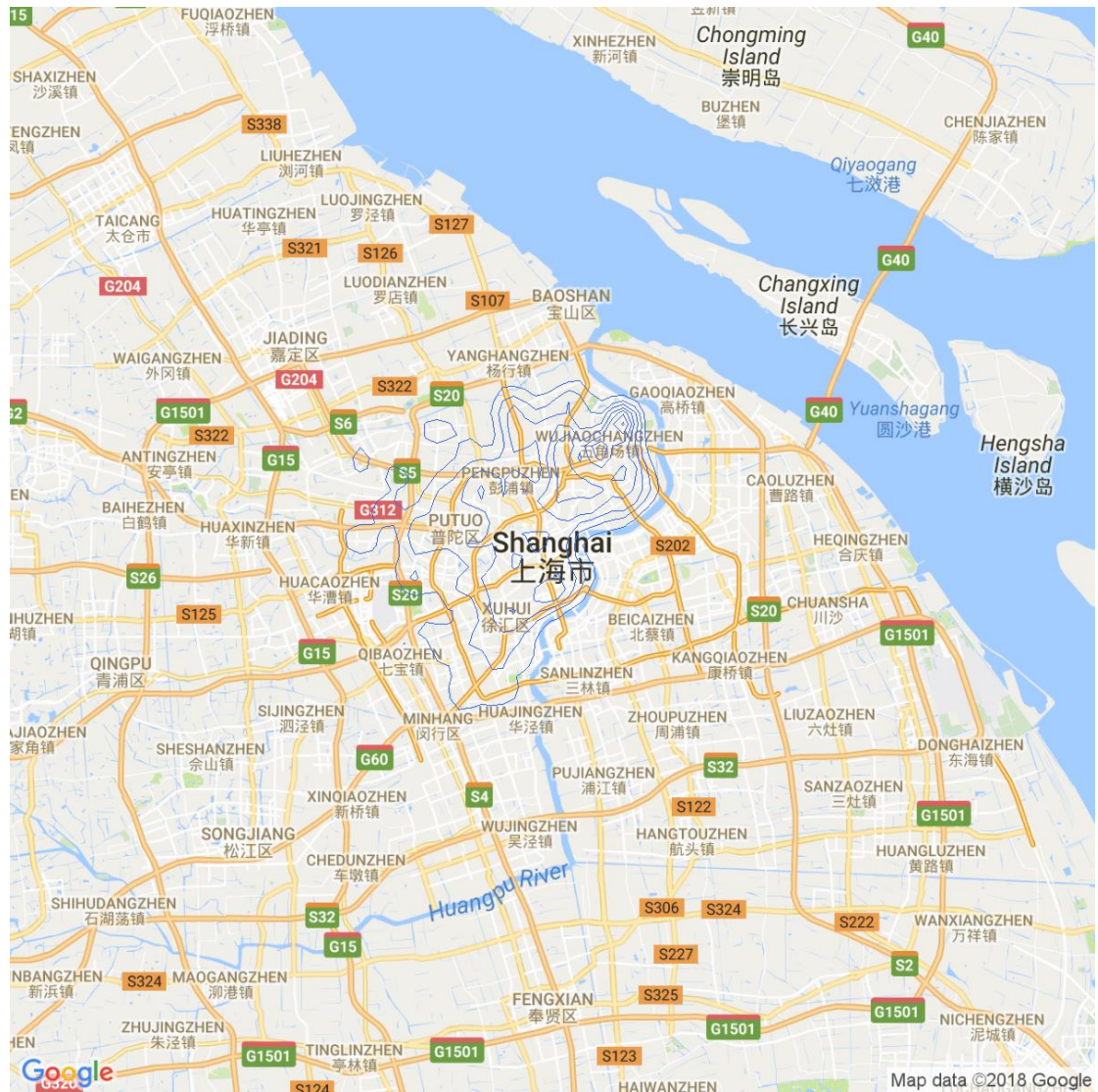


圖 4-16 上海摩拜單車還車密度曲線分佈圖 a

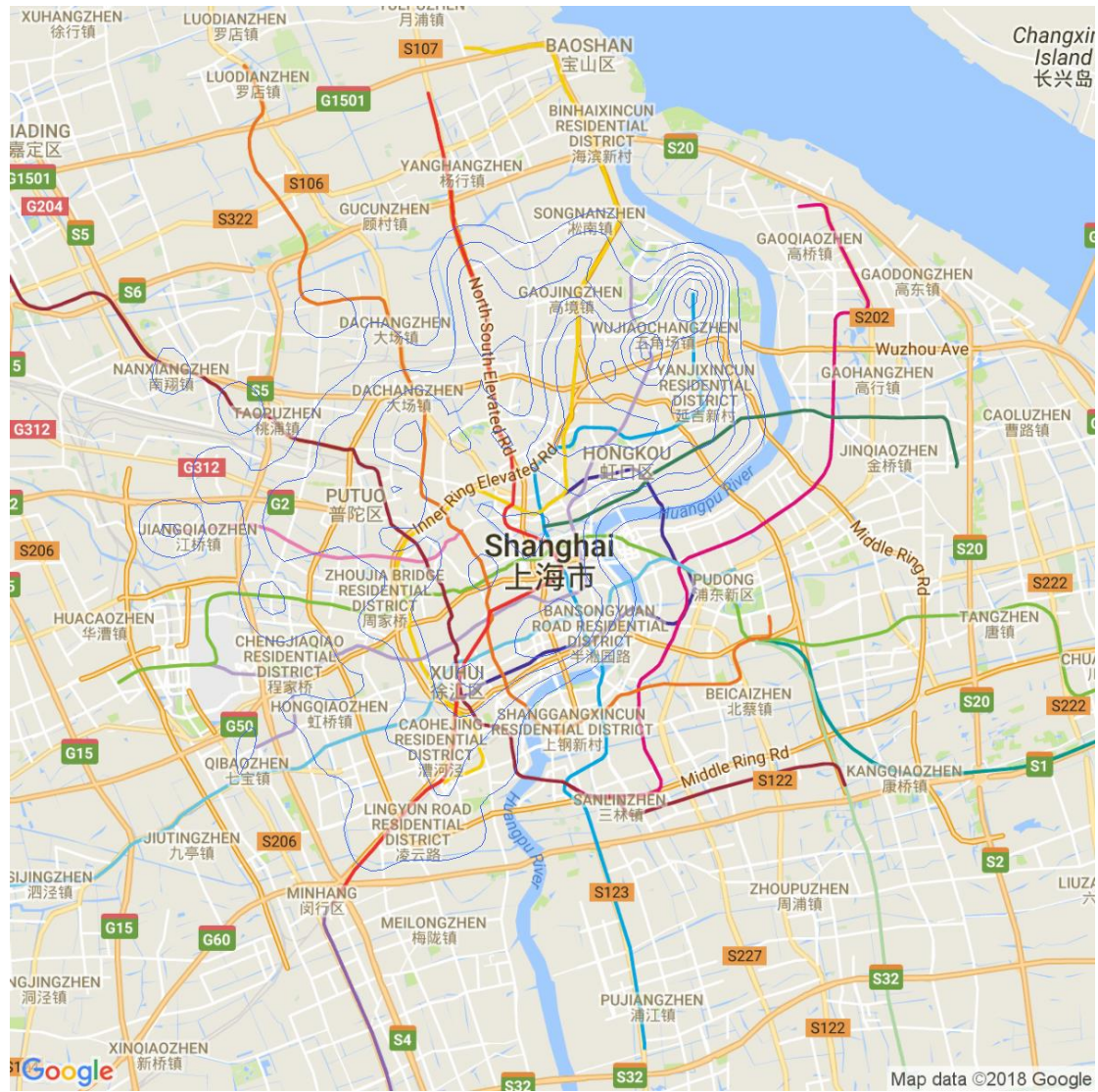


圖 4-17 上海摩拜單車還車密度曲線分佈圖 b

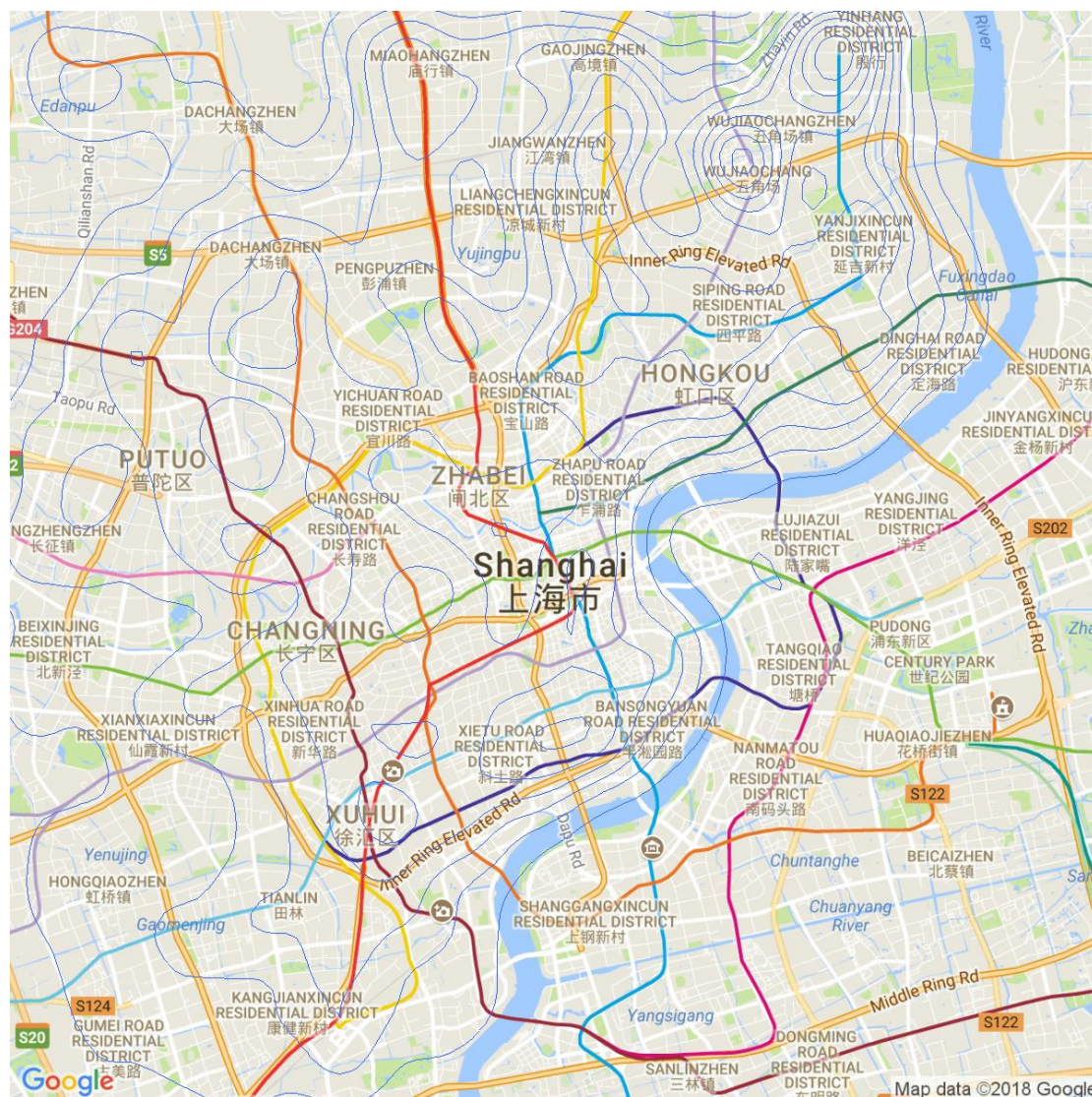


圖 4-18 上海摩拜單車還車密度曲線分佈圖 c

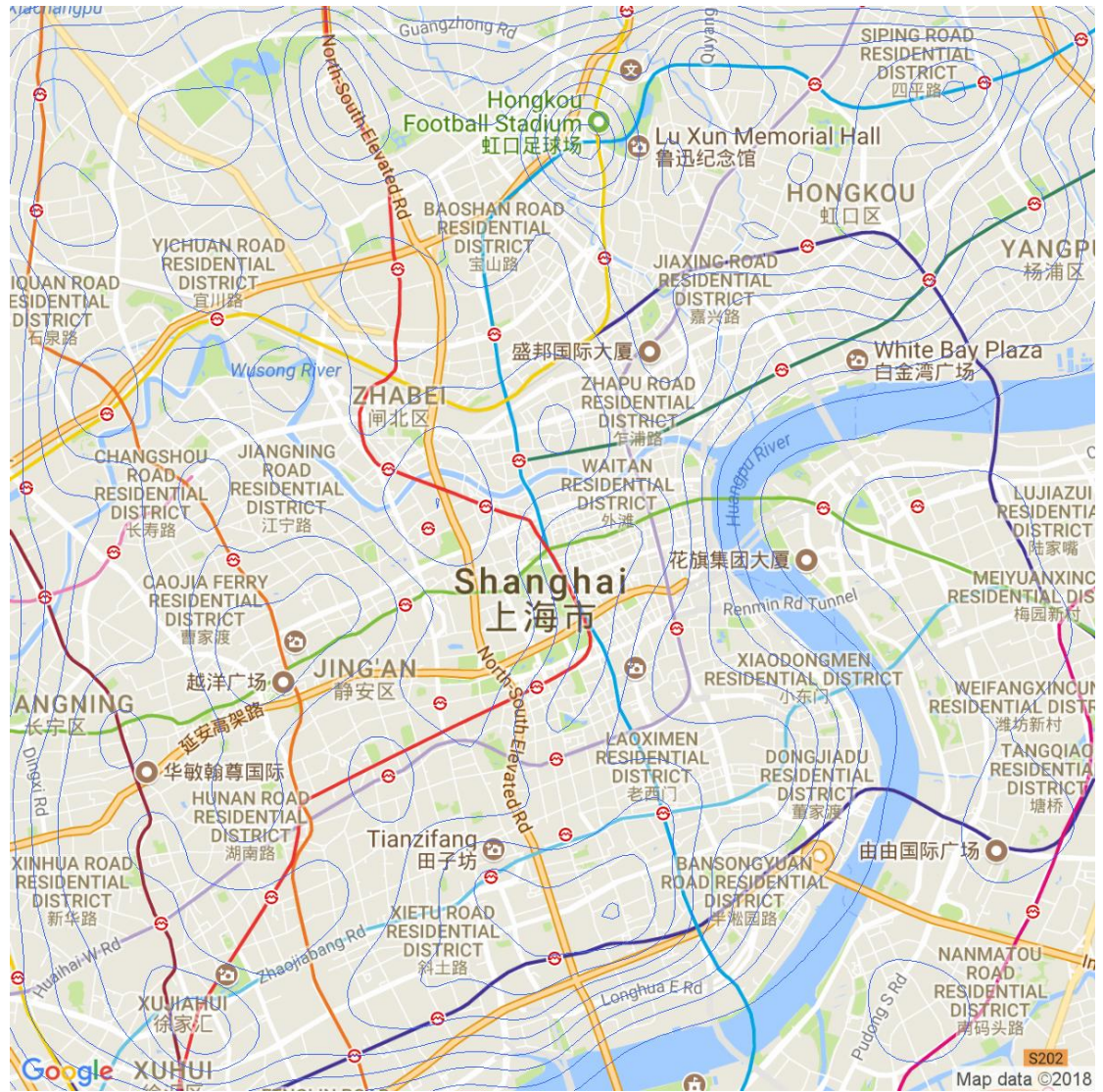


圖 4-19 上海摩拜單車還車密度曲線分佈圖 d

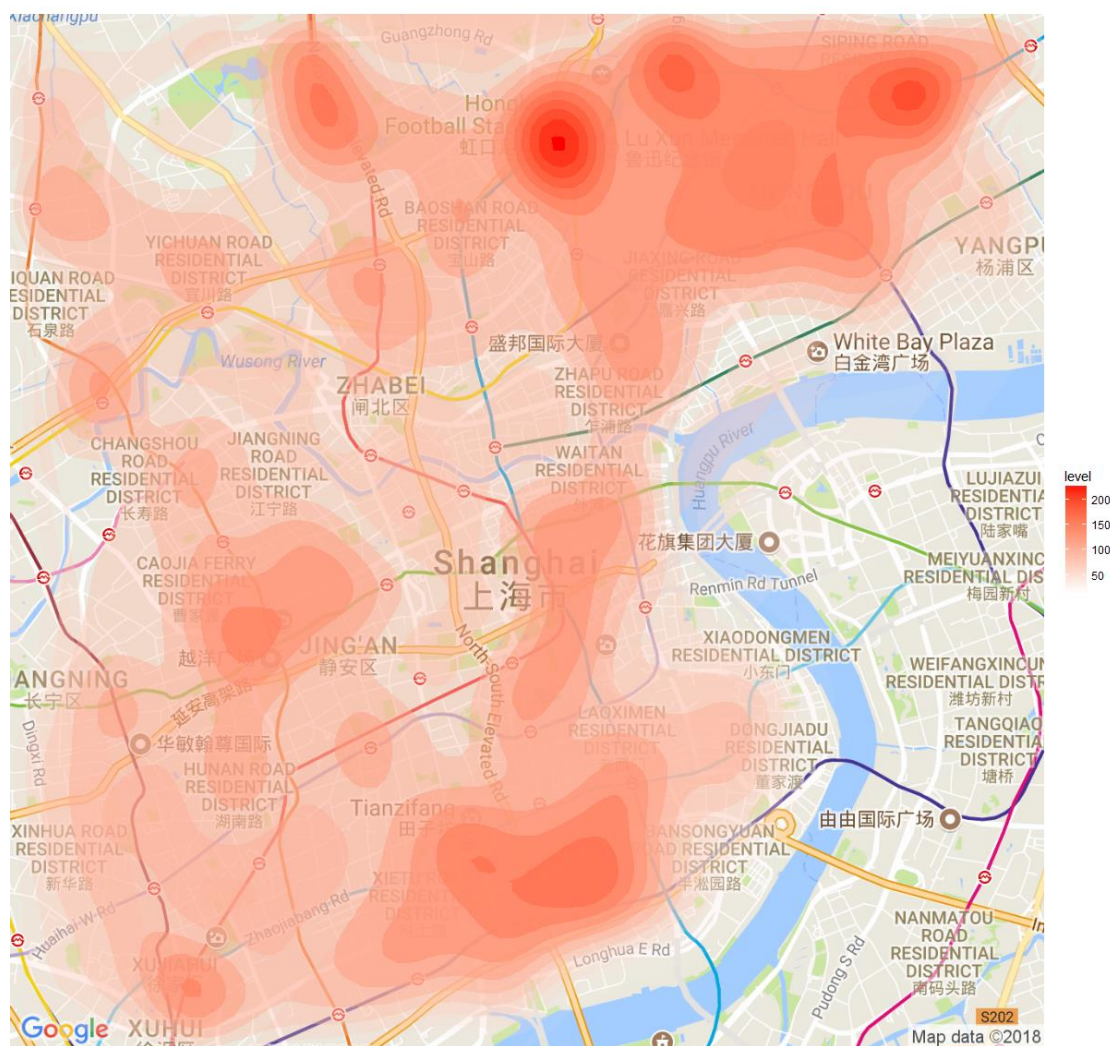


圖 4-20 上海摩拜單車還車密度曲線分佈圖

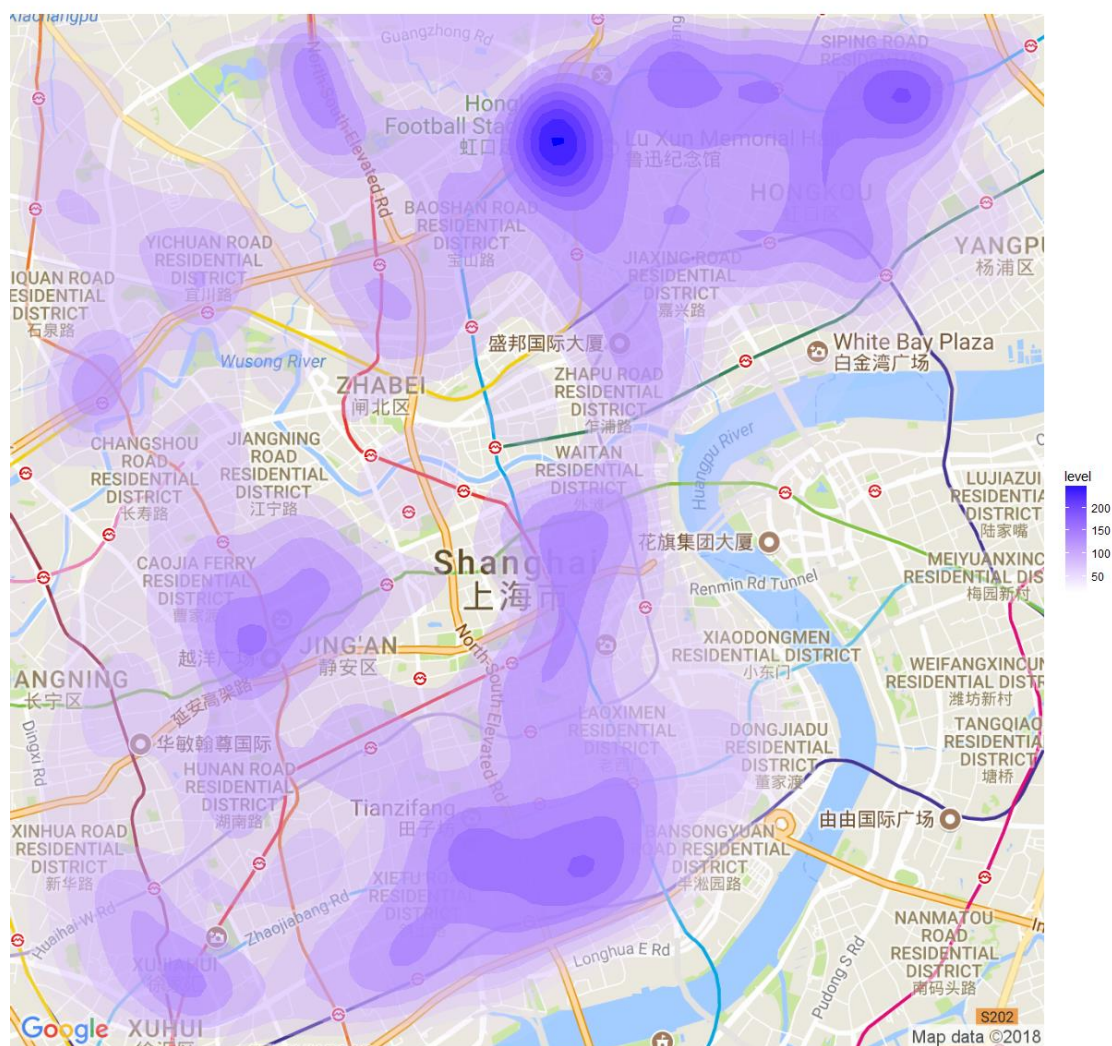


圖 4-21 上海摩拜單車取車密度曲線分佈圖

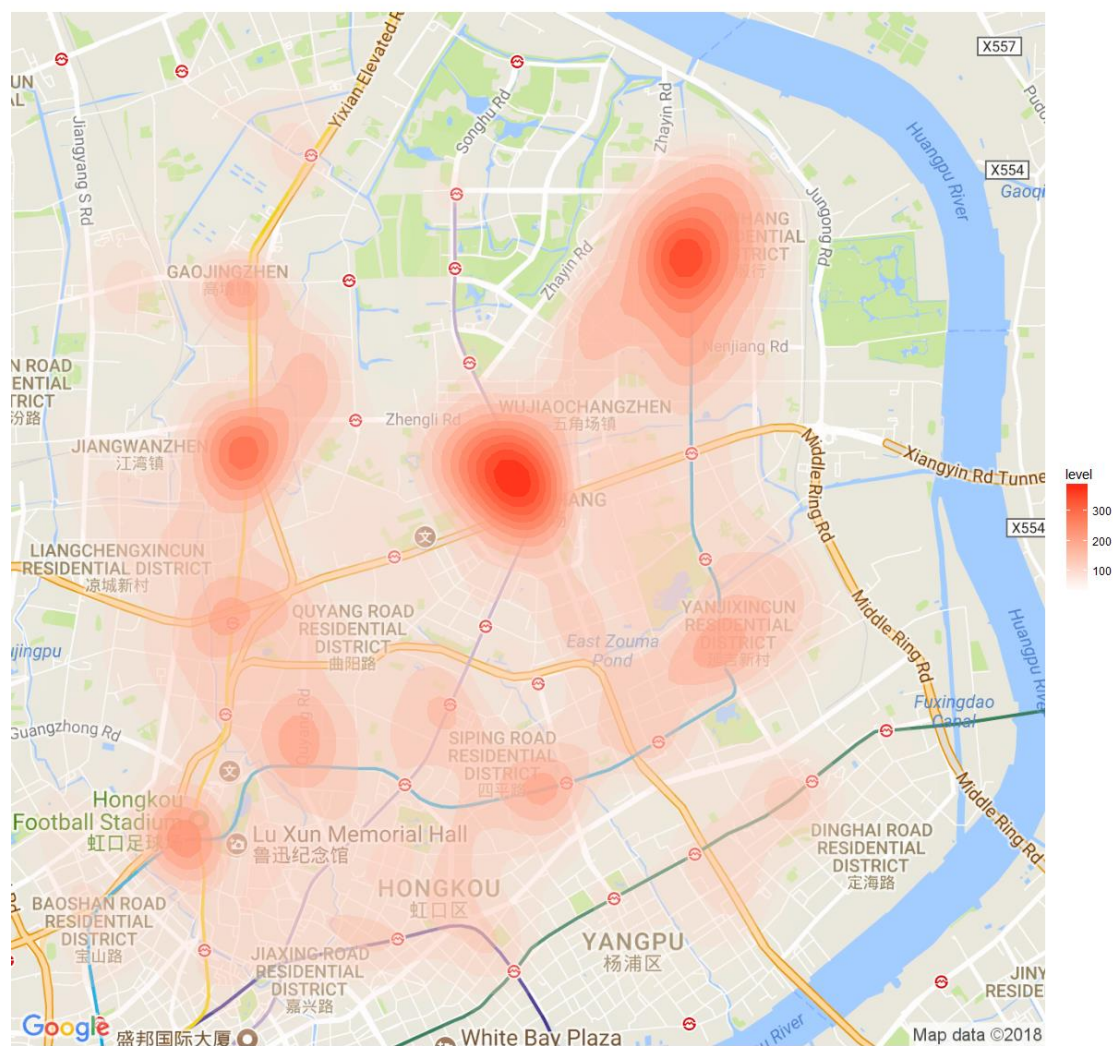


圖 4-22 上海楊浦區摩拜單車還車密度曲線分佈圖

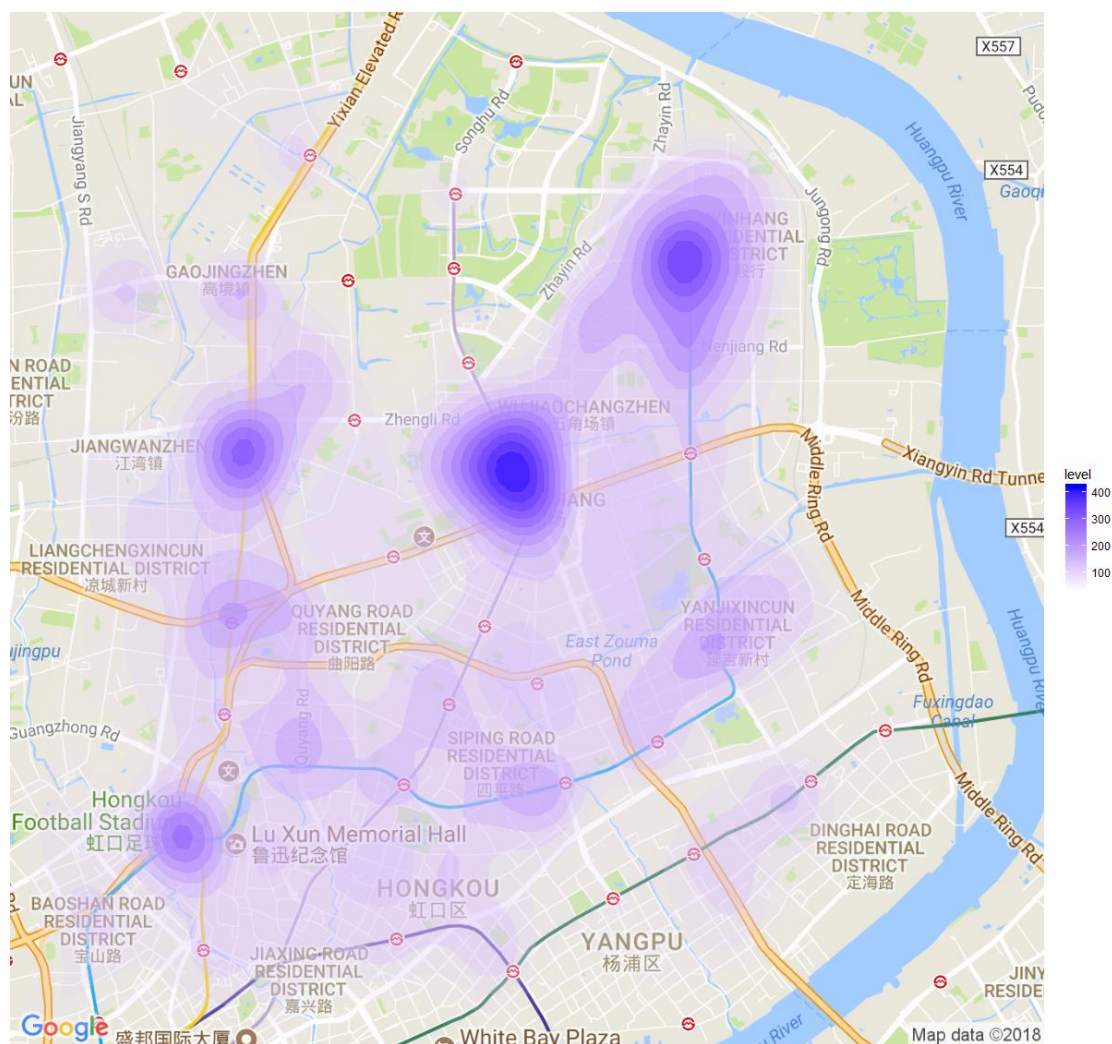


圖 4-23 上海楊浦區摩拜單車取車密度曲線分佈圖

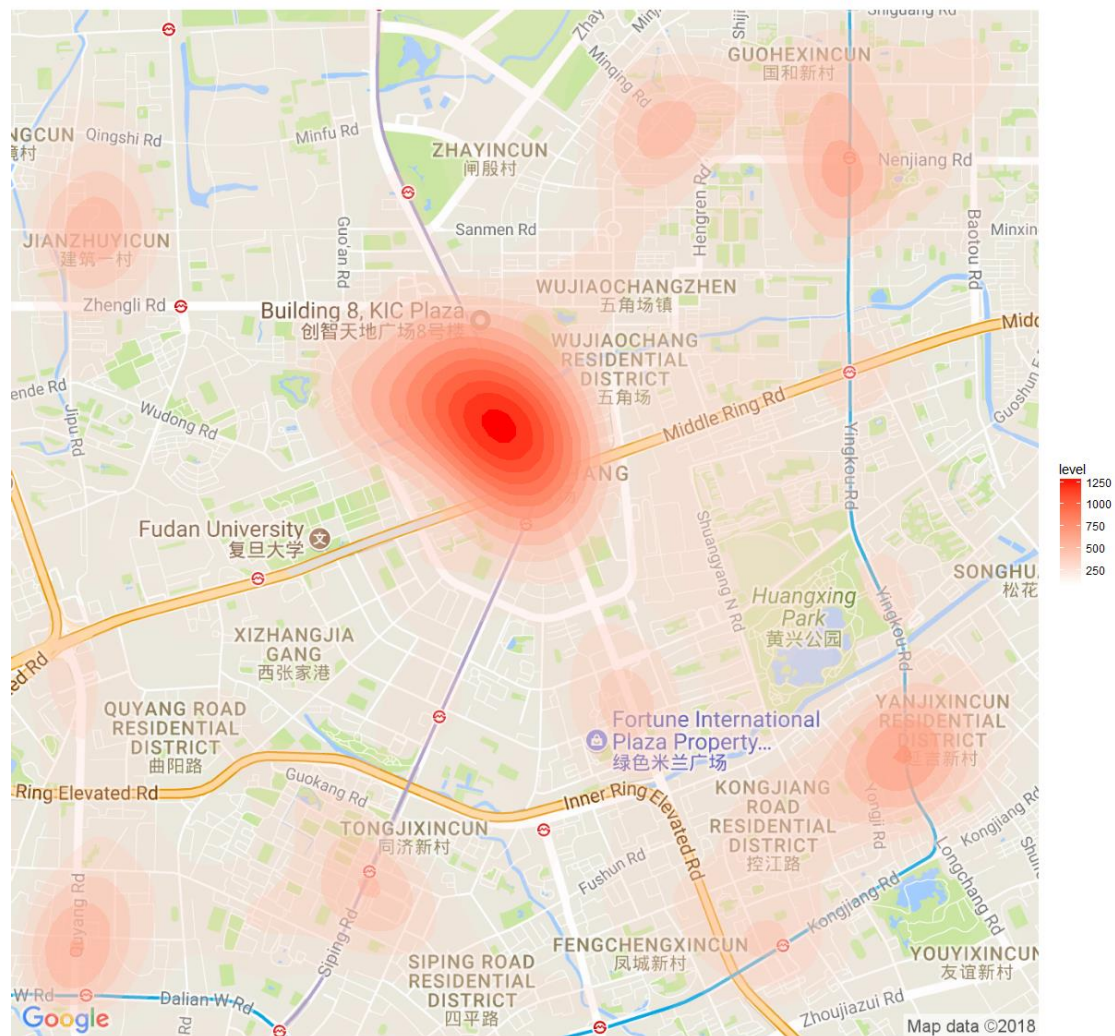


圖 4-24 上海五角場摩拜單車還車密度曲線分佈圖

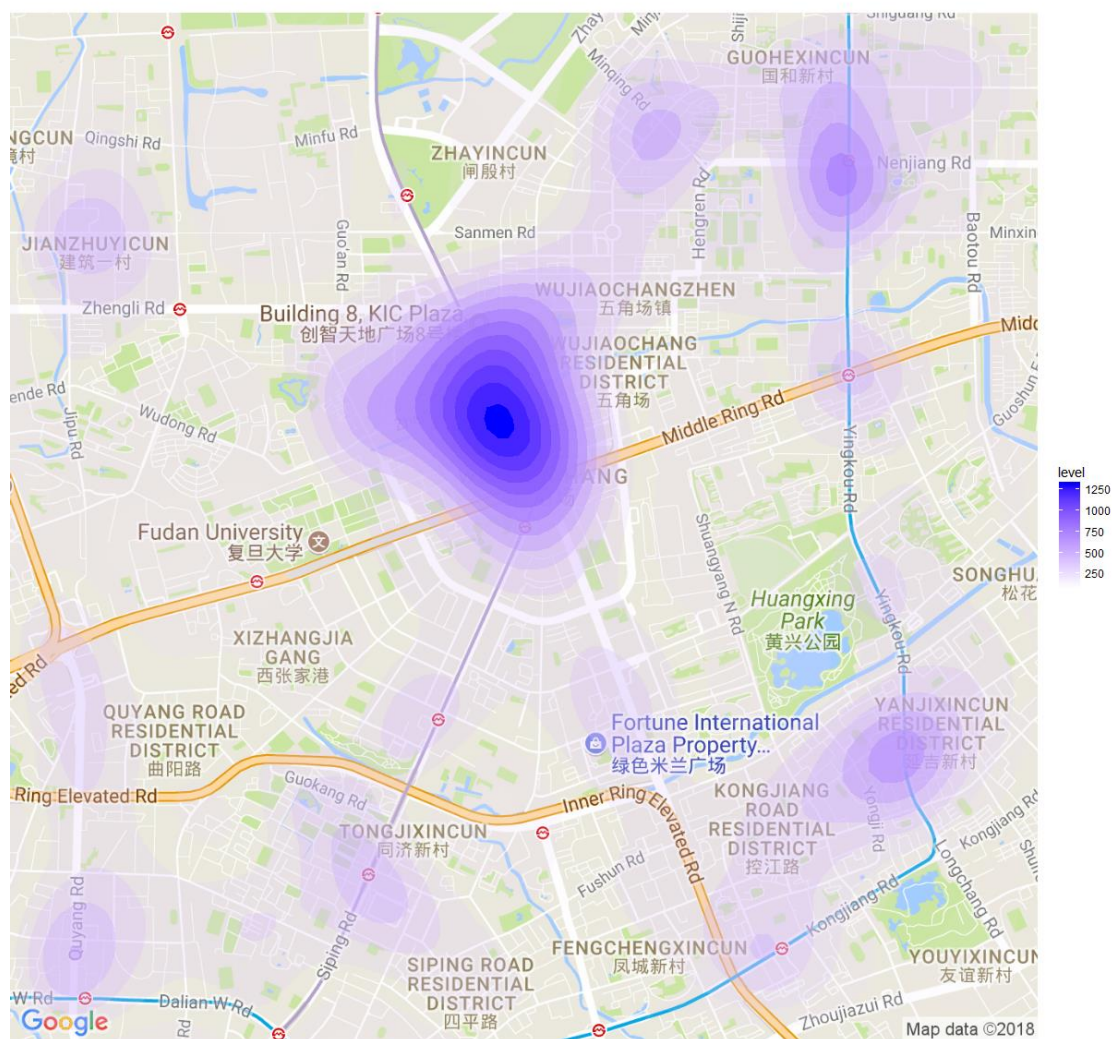


圖 4-25 上海五角場摩拜單車取車密度曲線分佈圖

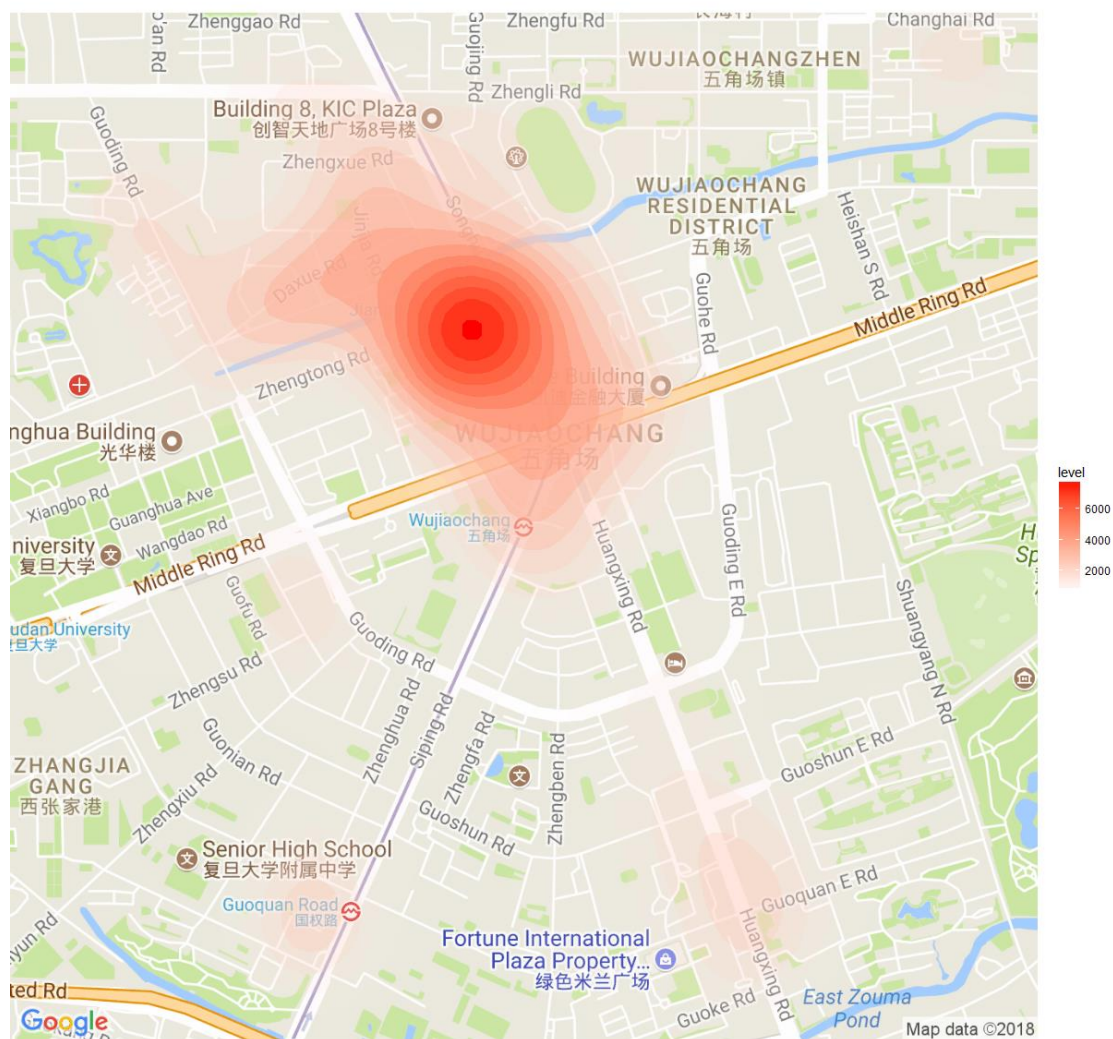


圖 4-26 上海五角場地鐵站摩拜單車還車密度曲線分佈圖

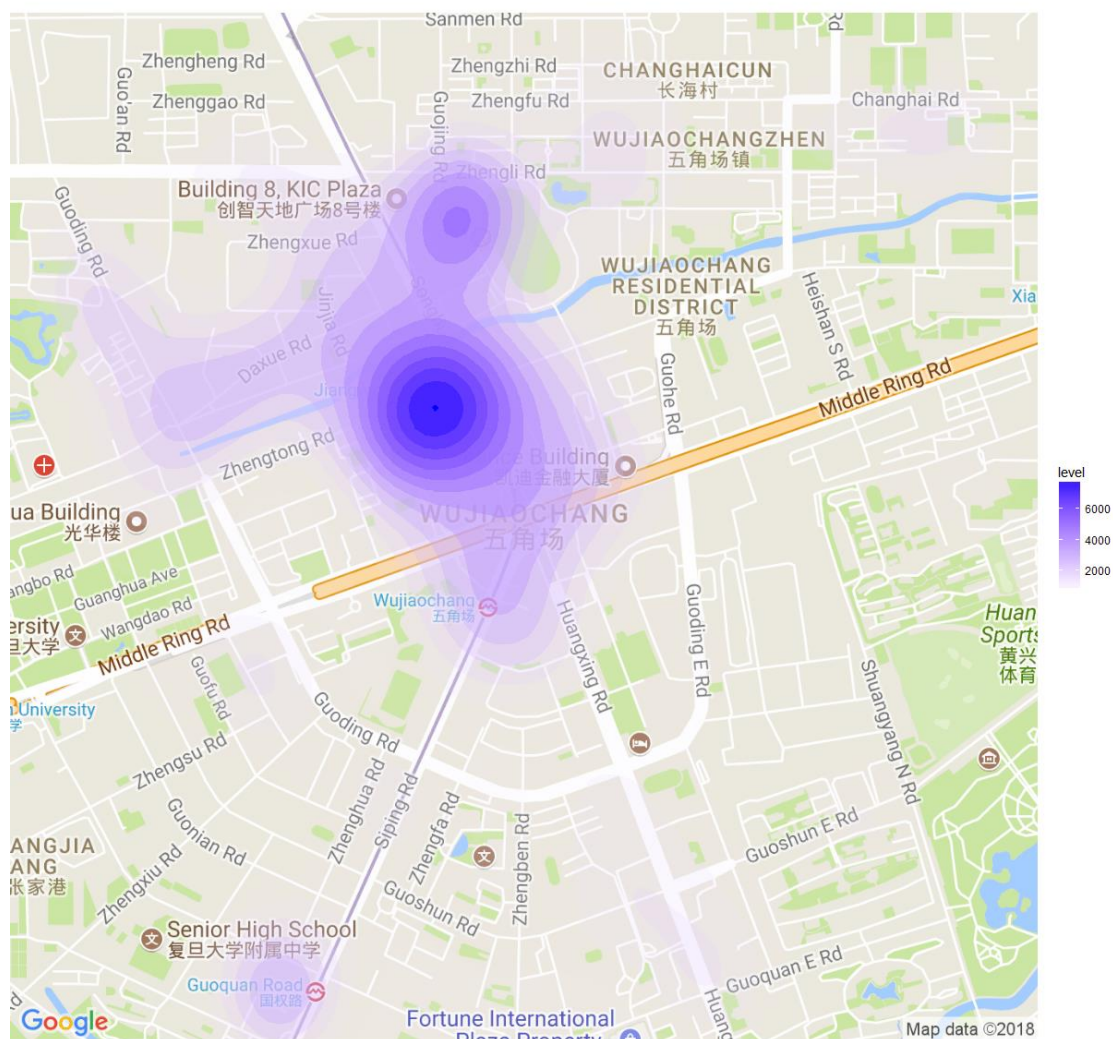


圖 4-27 上海五角場地鐵站摩拜單車取車密度曲線分佈圖

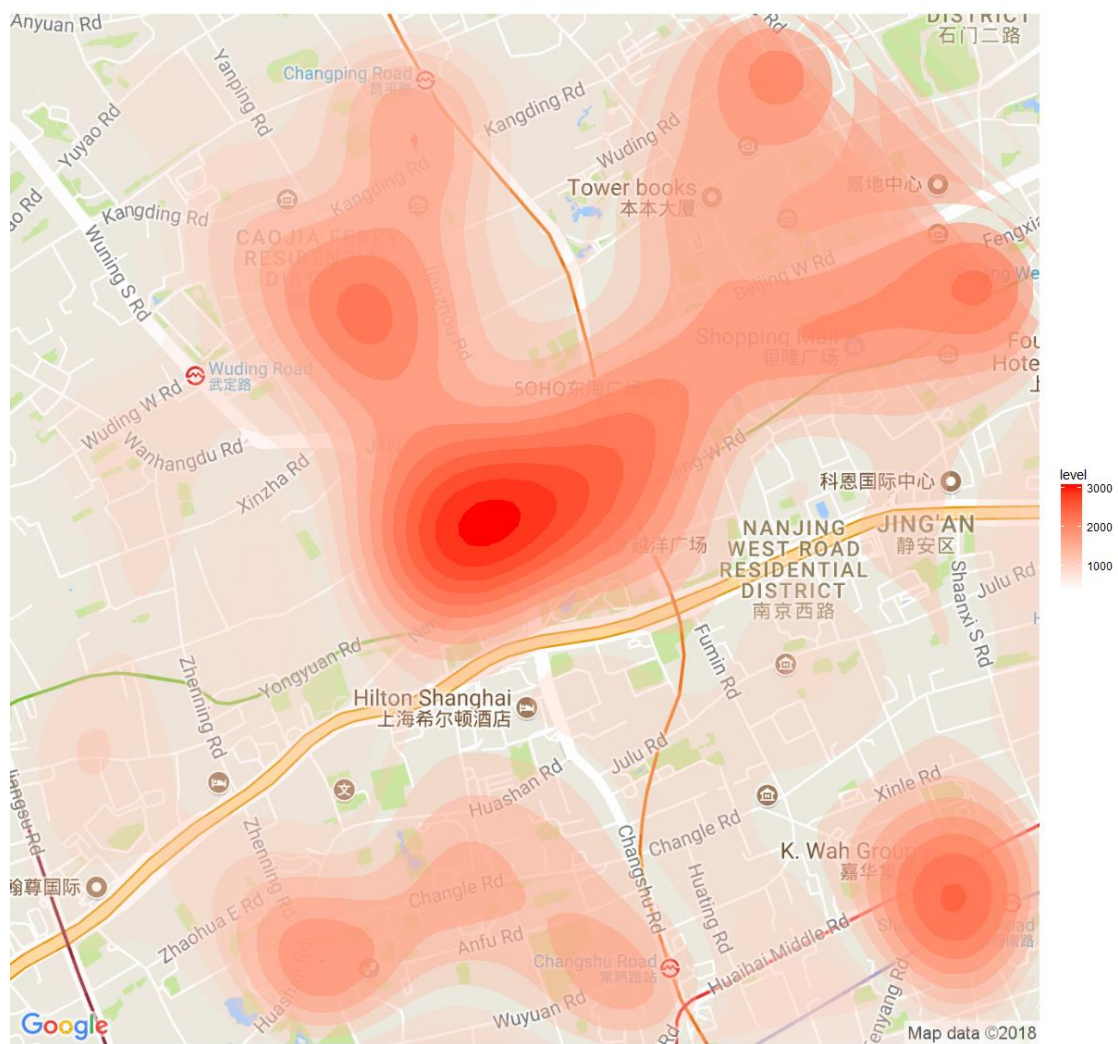


圖 4-28 上海靜安寺摩拜單車還車密度曲線分佈圖

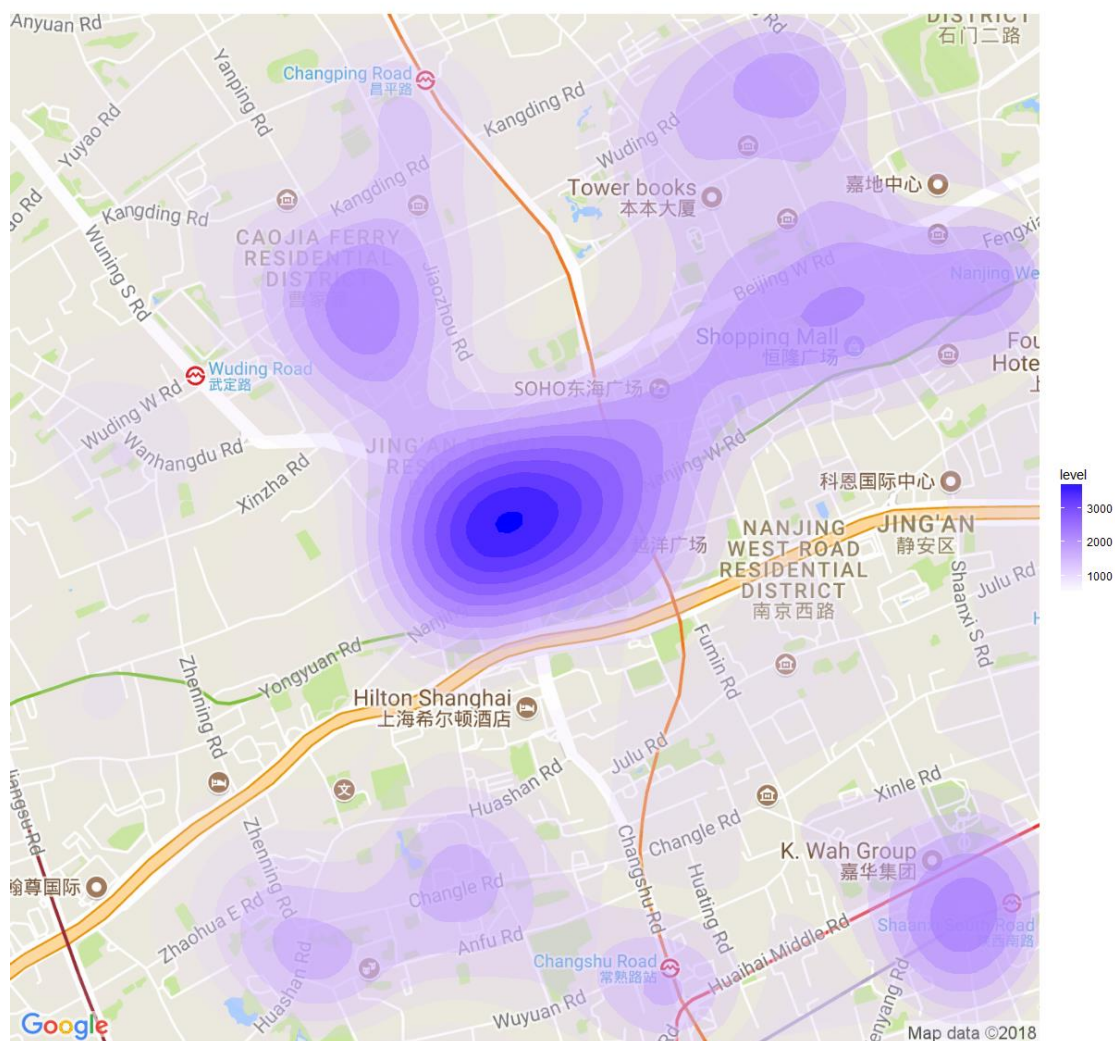


圖 4-29 上海靜安寺摩拜單車取車密度曲線分佈圖

經緯度方向550m 方格內摩拜單車平均存量約為1428輛。根據數據精確度只到經緯度千分位，即每100m 有一個點位，放大到550m*550m 方格內有25個點位，每一個點位上存在很多點的重合。根據上文採用的每個550m*550m 方格內單車平均存量為1428輛，25點位，每一點位平均存在57.12輛，此時用格子法劃分很難體現趨勢。

圖 4-30所示，顏色的深淺代表點位上單車輛數密度，越深色的表示密度越高，此時運用格子法劃分難以區分不同密度的點位元元：

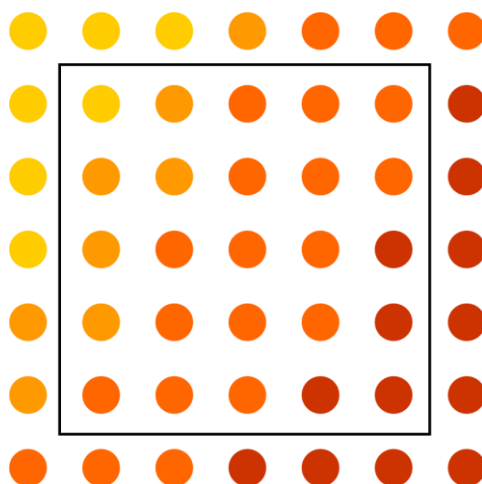


圖 4-30 550m *550m 範圍點位示意圖

在街道小尺度的情形下，本文建議運用密度分群法，根據每個點位上存在的實際單車量數，來作為分組，如圖 4-31。根據每個點位平均值57輛（對應550m*550m 方格內總數為平均存量1428輛）的數據，系統會給予合適的分群指標，比如每個點位上，存有範圍10-30個輛的點為一組，存有範圍30-50個點的分為一組，存有範圍50-70個點的分為一組等類似的分群方式。同時系統會忽略一些使得最終分群結果邊緣不連續的點位，來保證分群結果為較為圓滑的曲線。

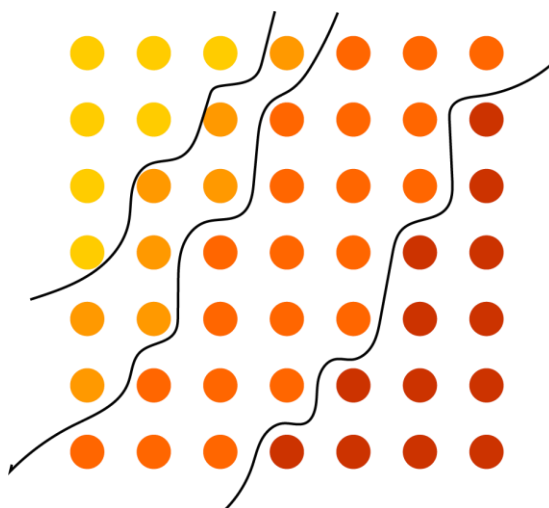


圖 4-31 根據點位元上存在的點密度分群示意圖

4.3 階段三：分群方法的探究及其實際運用

對於街道（靜安寺及其周邊）和城市（四環以內）兩種尺度，本文採用各自選擇其最佳分群方式的方法。對於較小的街道尺度採用”密度分群法”，對於較大的城市尺度採用”DBSCAN分群法”。以下就兩種尺度及其分區方法、調度機制具體說明：

4.3.1 應用密度分群法，處理較小的街道尺度，以上海靜安寺及其周邊為例

根據上海靜安寺2016.8每天上午6時至12時的還車密度，應用密度分群法得出還車密度分區(圖 4-32)。同樣根據取車密度，得出取車密度分區(圖 4-33)。為了保證資料的全面性，本文採用2016.8一個月的資料進行分析。

對比圖 4-32左與圖 4-33的情形，得知還車密度分區各級群和取車密度分區各級群在地圖上的範圍不同，導致同一運行軌跡的單車在兩種密度分區上處於不同的集群，圖 4-32為從1級到7級，圖 4-33為從1級到9級。在計算獎勵時採用分開計算各自獎勵的方式，最後相加得到最終的獎勵額。

I. 對於還車密度分區

建立獨立獎勵機制（圖 4-32），根據某輛單車取車點所在的取車區域，以及還車點所在的還車區域：

機器學習應用於共用單車流量分佈

若取車區域實行取車獎勵，還車區域實行還車獎勵，則把兩項獎勵的和作為獎勵額。

若取車區域實行取車獎勵，還車區域實行取車獎勵，則把取車區域的獎勵作為獎勵額。

若取車區域實行還車獎勵，還車區域實行取車獎勵，獎勵額為0。

若取車區域實行還車獎勵，還車區域實行還車獎勵，則把還車區域的獎勵作為獎勵額。

得到一個關於還車密度分區，最終的獎勵額度 A。

II. 對於取車密度分區

建立相同的獨立獎勵機制（圖 4-33）。根據同一輛單車取車點所在的取車區域，以及還車點所在的還車區域：

若取車區域實行取車獎勵，還車區域實行還車獎勵，獎勵額為0。

若取車區域實行取車獎勵，還車區域實行取車獎勵，則把還車區域的獎勵作為獎勵額。

若取車區域實行還車獎勵，還車區域實行取車獎勵，則把兩項獎勵的和作為獎勵額。

若取車區域實行還車獎勵，還車區域實行還車獎勵，則把取車區域的獎勵作為獎勵額。得到一個關於取車密度分區，最終的獎勵額度 B。

III. 最終的獎勵額度為 A 與 B 之和。

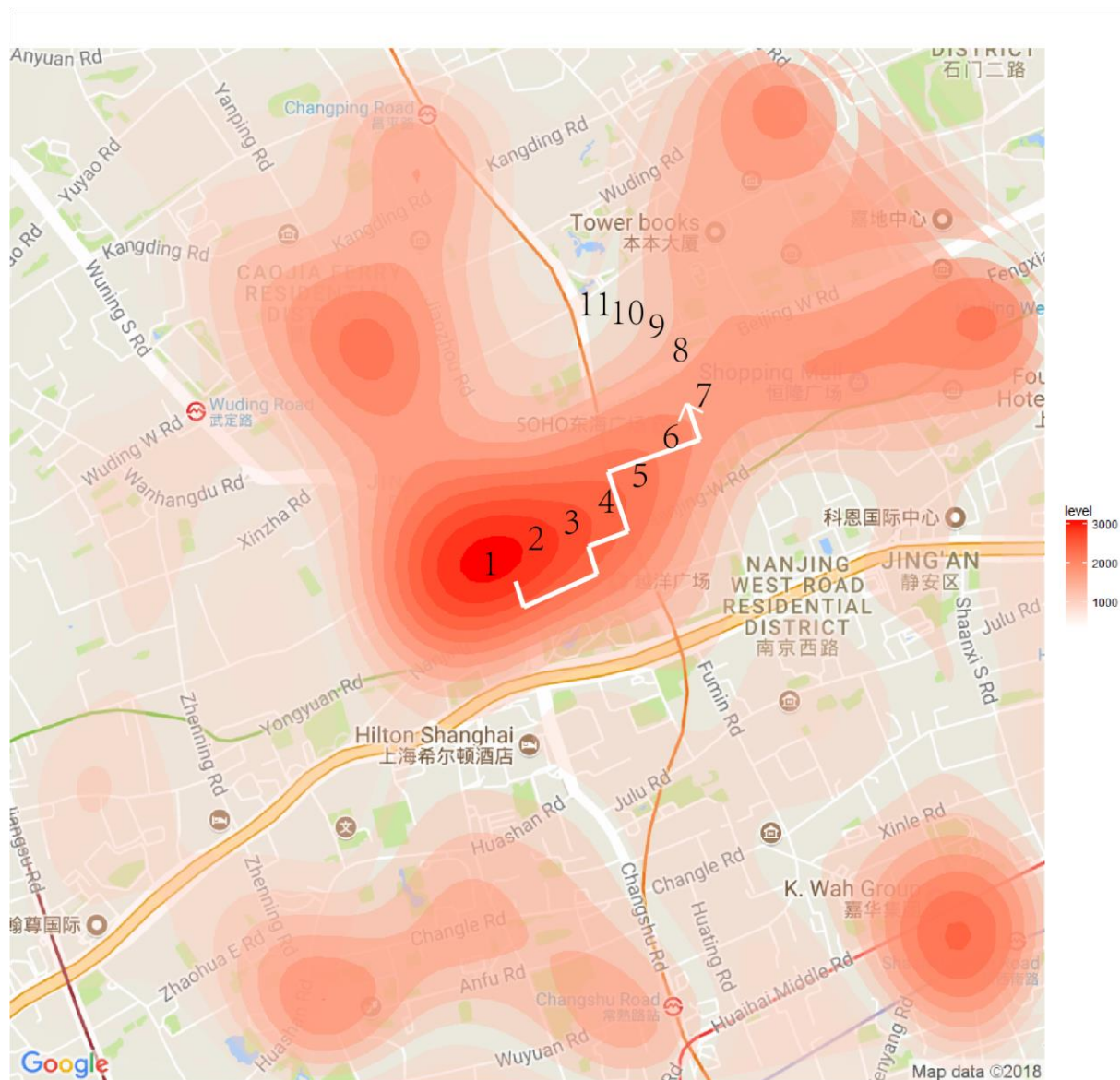


圖 4-32 上海靜安寺 2016.8 還車密度分區

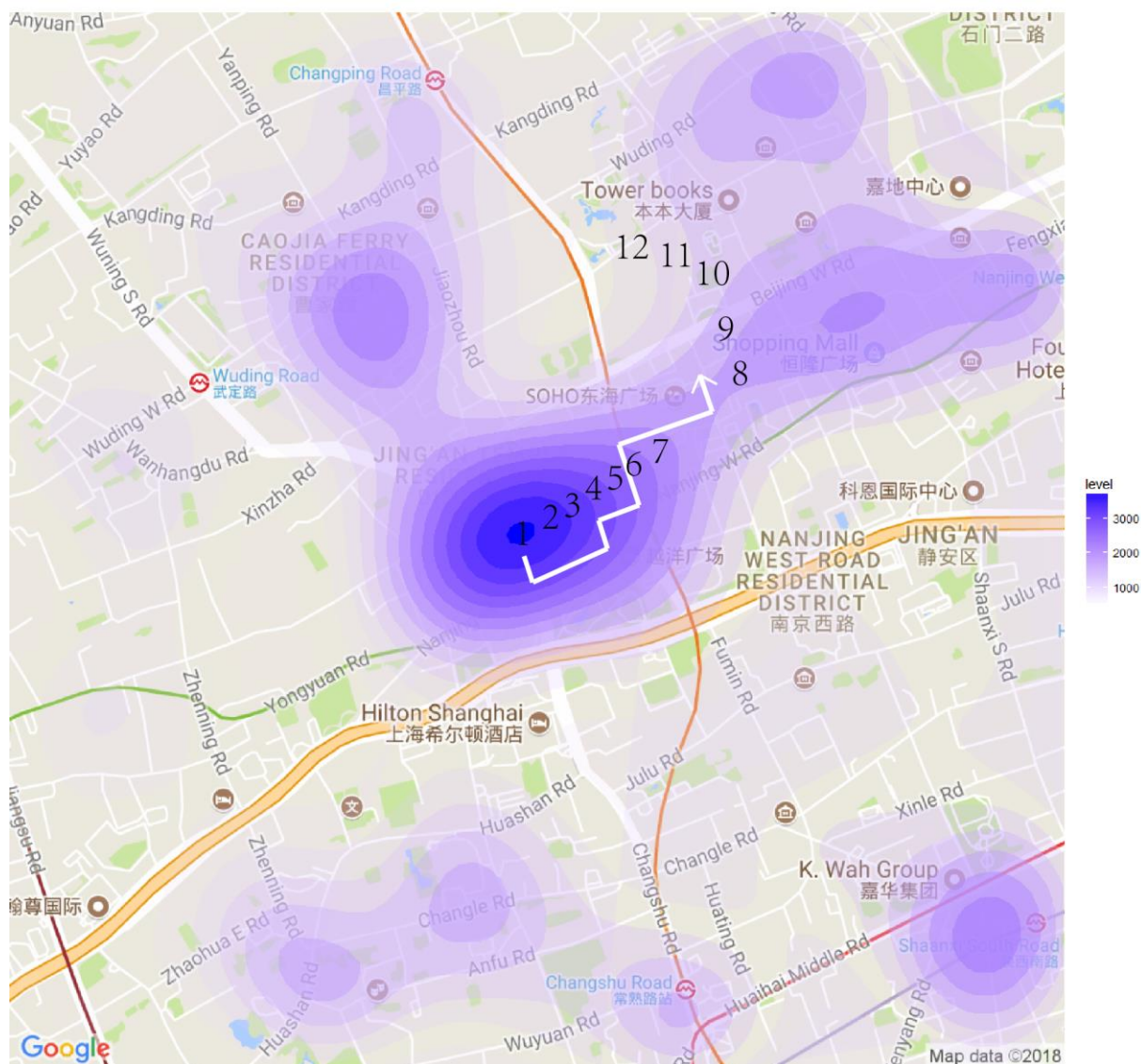


圖 4-33 上海靜安寺 2016.8 取車密度分區

▲ I 對於還車密度分區舉例

如圖 4-32計算標識出的結果，靜安寺中心1級區域還車密度約為3100（此區域內每個點位上存在點的個數約為 $3100 \div 25 = 124$ 個），外側邊緣區域還車密度約為800（此區域內每個點位上存在點的個數約為 $800 \div 25 = 32$ 個），得出中心區域到外側邊緣區域的密度差值約為 $3100 - 800 = 2300$ 。

如圖 4-32所示紅色深淺級數示意，系統給出的建議分群個數為11，將2300分為1到11個層級(levels)，得出每跨越相鄰兩個層級的還車密度變化約為 $1 \div 11 = 9.09\%$ ，轉換為數字為 $2300 \times 9.09\% = 210$ 輛。每個級數分區的閾值如表4-1：

表 4-1 還車分區各級數區域閾值

還車分區	1級	2級	3級	4級	5級	6級	7級	8級	9級	10級	11級
最高閾值	≥ 3100	2891	2681	2472	2263	2055	1845	1636	1427	1218	1009
最低閾值	2891	2681	2472	2263	2055	1845	1636	1427	1218	1009	≤ 800

圖4-33所示的還車密度分區情況為平衡狀態。

如圖4-34，若2016.9第一天上午6時至12時，第一輛單車的運動軌跡為從2級區域到7級區域：

2級區域的還車密度值為3100，高於最高閾值2891，則觸發“取車”獎勵，則該月的第二天上午6時至12時之間，2級區域上的點位上取車的車獲得總共 $3100 - 2891 = 209$ 的獎勵額度。第一輛取車的車獲得 $209 / 21 = 104.5$ 的獎勵額度。

7級區域的還車密度值為1600，低於最低閾值1636，則觸發“還車”獎勵，則該月的第二天上午6時至12時之間，2級區域上的點位上還車的車獲得總共 $1636-1600=36$ 的獎勵額度。第一輛還車的車獲得 $36/21=1.8$ 的獎勵額度。

即對於這輛單車來講，在還車密度分區獲得的獎勵為 $A=104.5+1.8=106.3$ 。

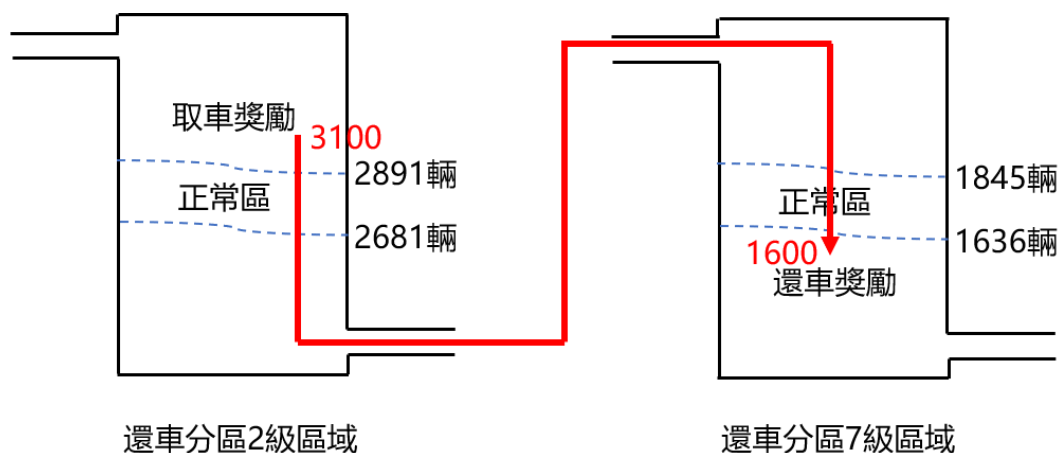


圖 4-34 單車於還車分區獎勵示意圖

▲ II 對於取車密度分區

如圖 4-33 計算標識出的結果可示，靜安寺中心1級區域取車密度約為3600（此區域內每個點位上存在點的個數約為 $3600 \div 25 = 144$ 個），外側邊緣區域還車密度約為900（此區域內每個點位上存在點的個數約為 $900 \div 25 = 36$ 個），得出中心區域到外側邊緣區域的密度差值約為 $3600 - 900 = 2700$ 。

如圖 4-33 所示紫色深淺級數示意，系統給出的建議分群個數為12，將2700分為1到12個層級(levels)，得出每跨越相鄰兩個層級的還車密度變化約為 $1 \div 12 = 8.3\%$ ，轉換為數字為 $2700 \times 8.3\% = 225$ 輛。每個級數分區的閾值如表4-2：

表 4-2 取車分區各級數區域閾值

機器學習應用於共用單車流量分佈

取車分區	1級	2級	3級	4級	5級	6級	7級	8級	9級	10級	11級	12級
最高閾值	≥3600	3375	3150	2925	2700	2475	2250	2025	1800	1575	1350	1125
最低閾值	3375	3150	2925	2700	2475	2250	2025	1800	1575	1350	1125	≤900

視圖4-33所示的取車密度分區情況為平衡狀態。

如圖4-35，若2016.9第一天上午6時至12時，同一輛單車的運動軌跡為從2級區域到9級區域：

2級區域的取車密度值為3000，低於最低閾值3150，則觸發“取車”獎勵，則該月的第二天上午6時至12時之間，2級區域上的點位上取車的車獲得總共 $3150-3000=150$ 的獎勵額度。第一輛取車的車獲得 $150/21=75$ 的獎勵額度。

9級區域的還車密度值為1500，低於最低閾值1575，則觸發“取車”獎勵，則該月的第二天上午6時至12時之間，2級區域上的點位上還車的車獲得總共 $1575-1500=75$ 的獎勵額度。第一輛還車的車則不接受這部分的取車獎勵，即獎勵額為0。

即對於同一輛單車來講，在還車密度分區獲得的獎勵為 $B=75+0=75$ 。

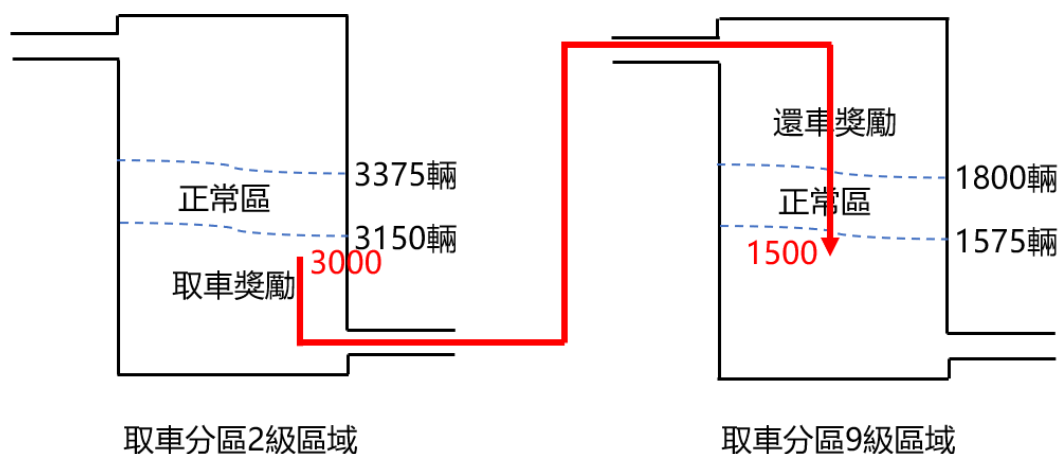


圖 4-35 單車於取車分區獎勵示意圖

4.3.2 應用 DBSCAN 分群法，處理較大的城市尺度，以上海四

環及其內部為例

出於數據完整度的考量，本節對於紐約市計程車上客點位置的資料 (Citibike, 2017)，分別運用 K-平均演算分群法與 DBSCAN 分群法，進行分群，通過比較分群的結果，考量兩種分群法的優劣。

首先採用 K-平均演算分群法，取 $K=5$ 的情況下做分群。如圖 4-36 可示，劃分得到 5 個不同顏色分群的結果，把結果放入實際地圖中，如圖 4-37、圖 4-38 所示，為參數設定相同時分別進行兩次運算得的結果，可見其分群結果的隨機性，驗證了 K-平均演算分群法具有不能區分形狀較任意的類簇的缺點。K-平均演算分群法首先需要確定劃分 K 個部分，而不適用於對分佈形狀不規則的交通資料做密度分群。

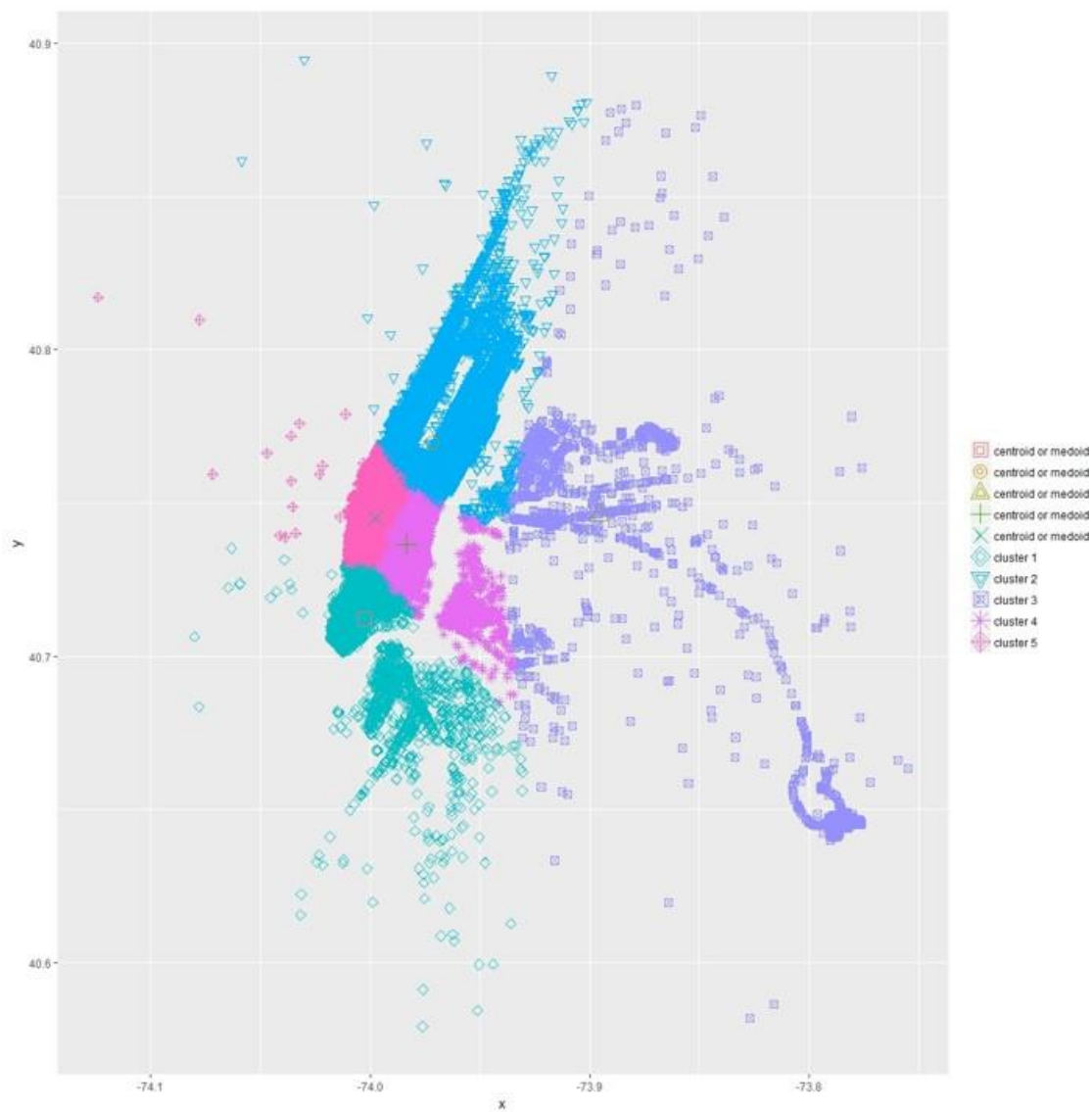


圖 4-36 K-平均演算分群法對於紐約計程車上客點位分析

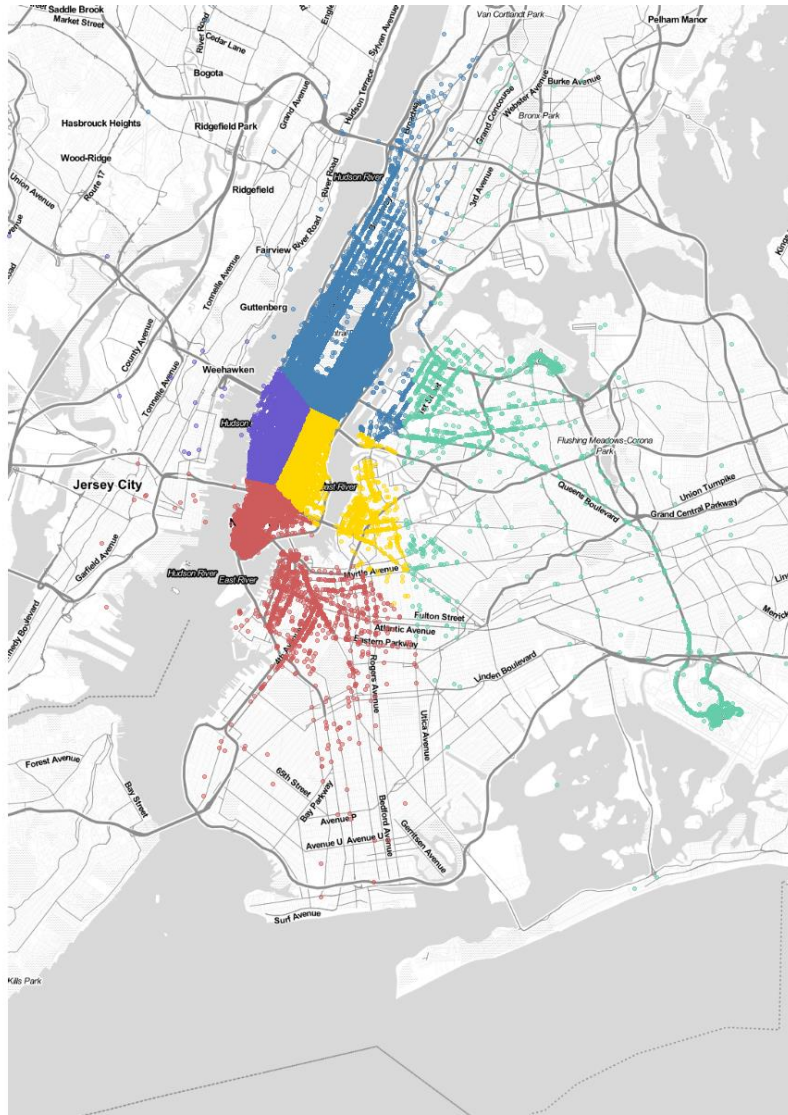


圖 4-37 K-平均演算分群法對於紐約計程車上客點位地圖

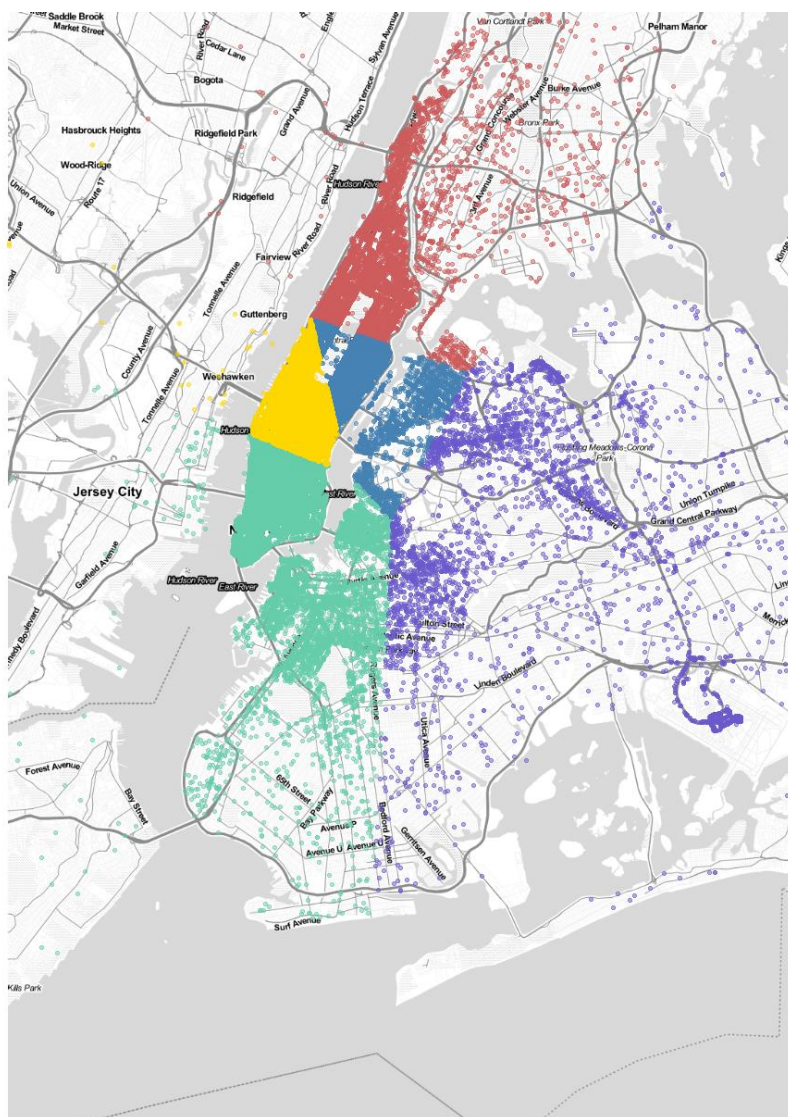


圖 4-38 參數設定相同 K-平均演算分群法對於紐約計程車上客點位地圖

圖 4-39是對於同一筆資料，運用 DBSCAN 分群法，取掃描半徑0.5，最少點數2000做分群，劃分得到8個彩色群組，都滿足掃描範圍內最少點數大於基礎值的要求，剩餘黑色點未達要求，為噪點。驗證了其兩個優點：（一）由於最少點數參數，單一連接現象（single-link effect，不同聚類以一點或極幼的線相連而被當成一個聚類）能有效地被避免。（二）能夠分辨噪點。可見 DBSCAN 分群法適用於對大尺度、數據分佈較為複雜的交通資料作密度分群。

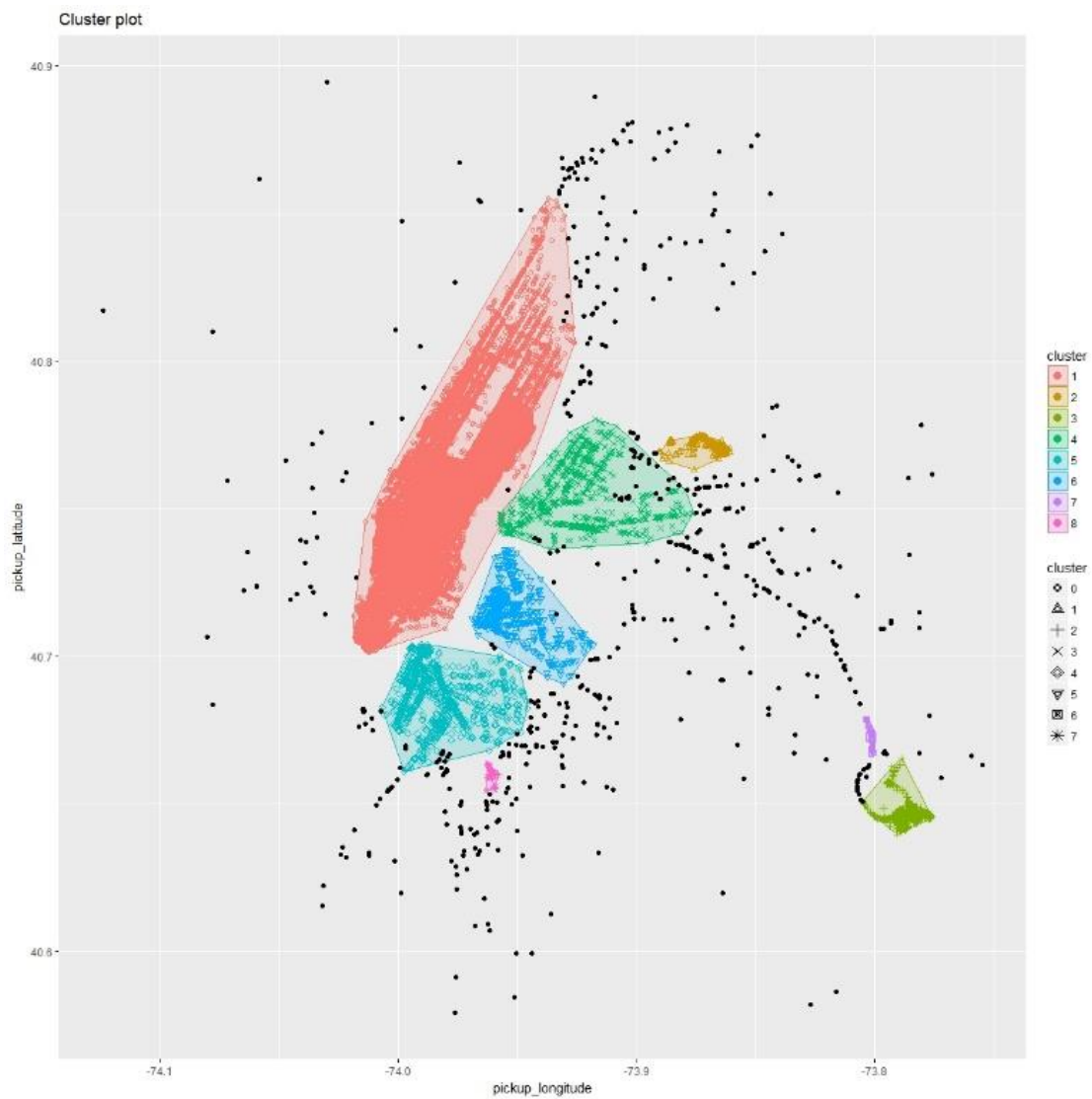


圖 4-39 DBSCAN 分群法對於紐約計程車上客點位分析

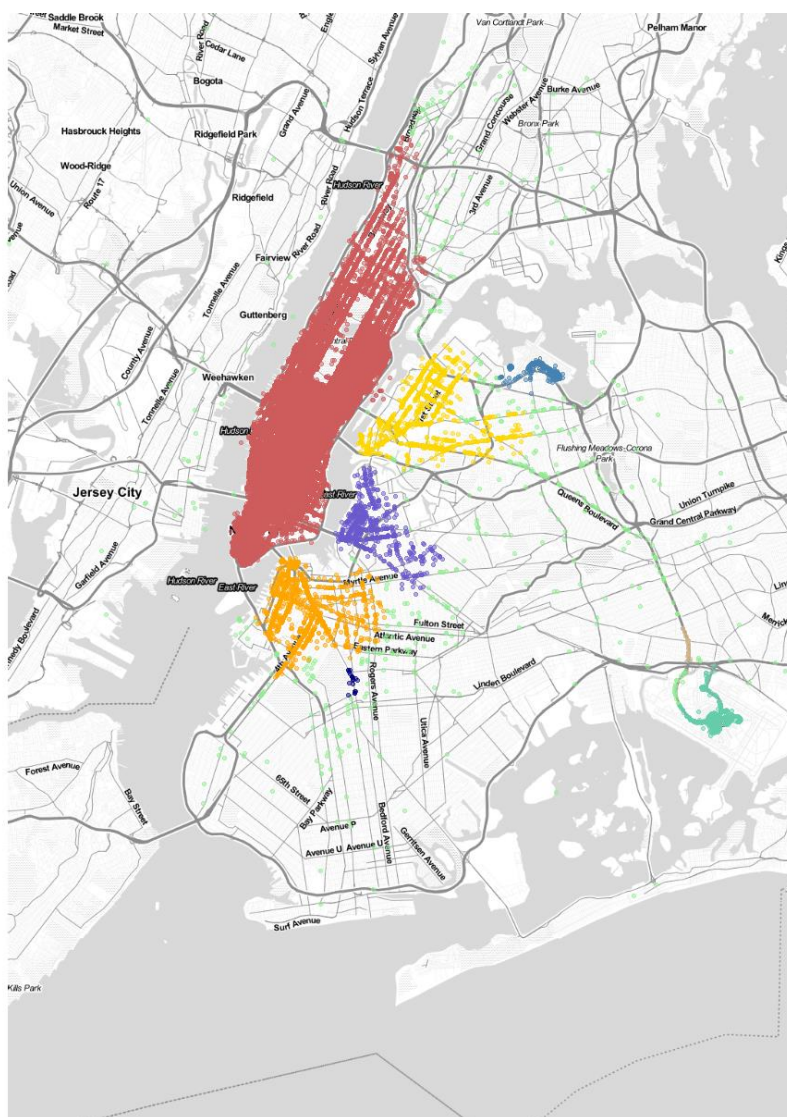


圖 4-40 DBSCAN 分群法對於紐約計程車上客點位地圖

所以處理較大的城市尺度，以上海四環（ 121.75°E - 121.125°E ， 30.95°

N - 31.45°N ）為例，本研究採用 DBSCAN 分群法：

根據在中心點位，使用者在尋找到可用單車前，步行距離應以不超過最適步行距離為限，即取步行半徑為700m的圓形範圍。

如圖 4-41所示，標識出圓形範圍與550m*550m 格子之間的關係。設定掃描半徑為0.7 (700m)，此時對於每個中心點範圍內的點位總共為176個。

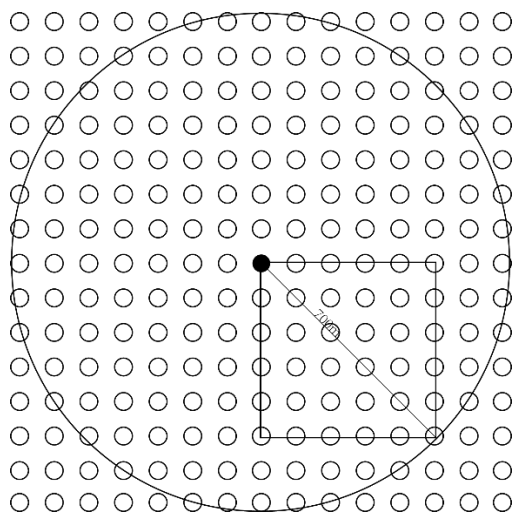


圖 4-41 掃描半徑 700m 範圍示意圖

根據本章第二節對於點和點位關係的探究，得知圖 4-41所示每個點位上存在多個點的重合，所以不止要考量範圍內最少點數，同時需要考量範圍內過半點位上的點數滿足 \geq 平均值，以避免出現範圍內最少點數達到標準、但點數的分佈情況不夠均勻的情形。

關於使用者的角度，找到可用單車，則意味著應根據圓形範圍內的單車供給量要充足。根據前文描述的將單位時間內單車的還車量視作單車的供給，本節對上海四環範圍內2016.8單車的供給分佈點位進行 DBSCAN 分群。

設定三種不同的最少點數，以篩選在最少點數呈增加狀態的時候仍滿足條件的區域。關於最少點數值的設定，考量每個點位上存在的點個數平均值。根據前文根據實際數據的估算，取

機器學習應用於共用單車流量分佈

每個點位上存在的點個數平均值57輛（對應550m*550m 方格內平均存量1428輛）。以及根據靜安寺2016.8還車分區1級區域的最高閾值3100輛，計算出的此區域內每個點位上存在點的個數約為 $3100 \div 25 = 124$ 輛。同樣根據同分區11級區域的最低閾值800輛，計算出的此區域內每個點位上存在點的個數約為 $800 \div 25 = 32$ 輛。區域內最少點數以平均每個點位32, 57, 124為基礎計算，對應的值乘上點位數176，即：

設定最少點數值 $a = 32 \times 176 = 5632$ ，且同時176個點位中有88個點位存在點的個數 ≥ 32 ；結果如圖 4-42。

設定最少點數值 $b = 57 \times 176 = 10032$ ，且同時176個點位中有88個點位存在點的個數 ≥ 57 ；結果如圖 4-43。

設定最少點數值 $c = 124 \times 176 = 21824$ ，且同時176個點位中有88個點位存在點的個數 ≥ 124 ；結果如圖 4-44。

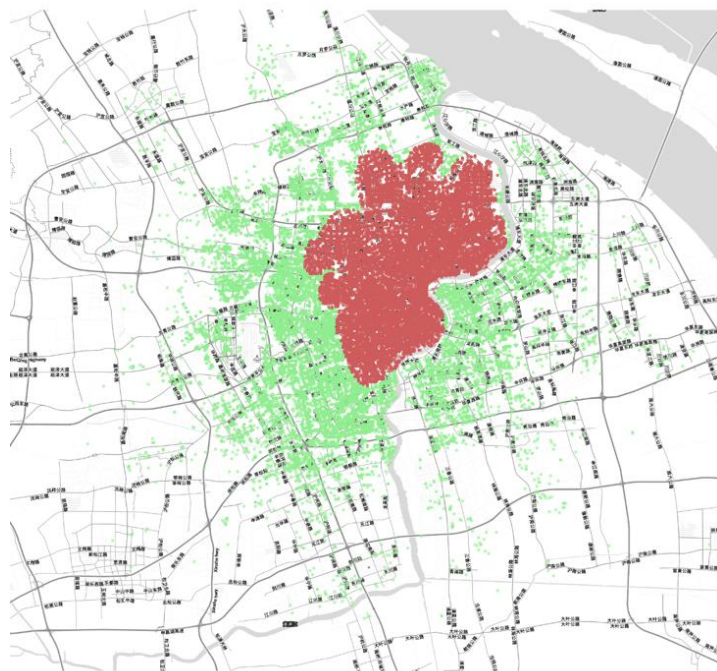


圖 4-42 還車點位，每個點位上有大於 32 個點單位元元面積的密度標示

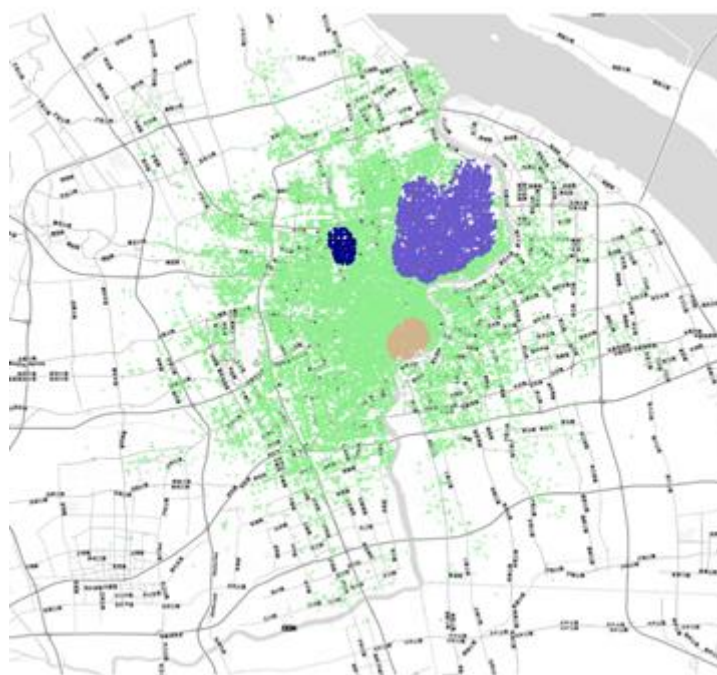


圖 4-43 還車點位，每個點位上有大於 57 個點單位元元面積的密度標示

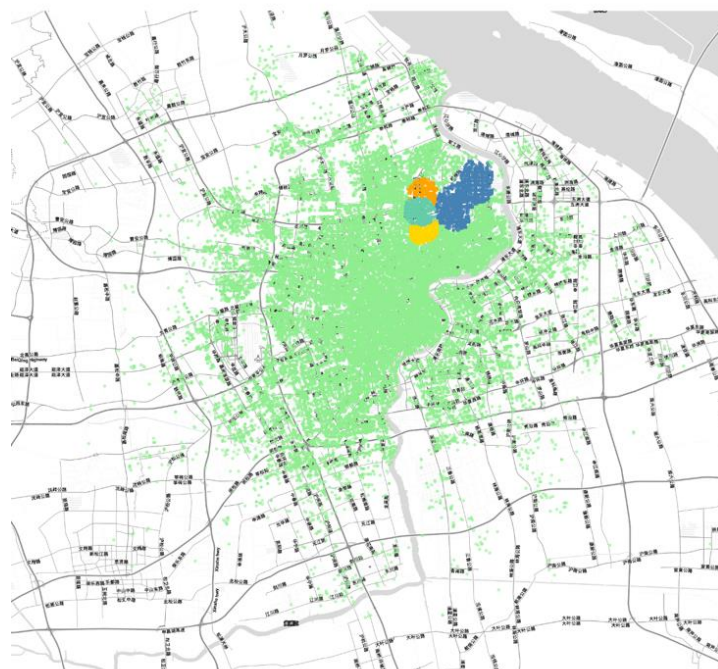


圖 4-44 還車點位，每個點位上有大於 124 個點單位元元面積的密度標示

可見在城市尺度上，DBSCAN 分群法能夠讓決策者及時掌控哪些區域出現了單車囤積的現象，提前安排公司的運輸卡車到達這些區域，及時進行強制調度。此法用於預防集中式的單車囤積困境，抉擇者可以根據現有運營的單車密度分佈，設定不同的最少點數值，以此比較單車在不同密度值下呈現的分佈狀態。以此做參考，來決定下一階段新投入運營單車的數量和投放地點。

4.4 階段四：疏散，調度策略總結

街道尺度實施使用者獎勵機制，根據還車密度分區、取車密度分區，各自建立獨立的獎勵系統，在計算獎勵時採用分開計算各自獎勵的方式，最終相加得到最終的獎勵額度。決策者根據某一時段內所有的數據知曉獎勵額度的總和，計算每一輛單車得到總獎勵的百分比數額。再根據節省下的原本人力運輸調度的費用，得出每輛單車得到的具體獎勵元數。

城市尺度由決策者調度，通過隨時設定最少點數，觀察區域性囤積的現象，做出及時的卡車運輸調度策略和新的運營、投放單車策略。

4.5 小結

本研究採用 R 語言、並使用” RStudio” 軟體作為編寫 R 語言程式的整合開發環境。作為視覺化的工具，當腳本將數據資料演變為圖像，使用者與決策者能夠直觀地瞭解單車點位元元與密度的實際分佈。

而機器學習的聚類分析 (cluster analysis) 則採行三種方法以進行測試、與比較分析。三種方法分別是：K-平均演算分群法 (K-means)、密度分群法、以及 DBSCAN 分群法 (Density-based spatial clustering of applications with noise)。

在實際測試運用，分析比較的過程中，分別以靜安寺及其周邊、上海四環 (121.75°E-121.125°E, 30.95°N-31.45°N) 及其內部為兩種不同空間尺度的研究範圍，探討在不同空間尺度下，最佳的分區方式。以下就這兩種尺度為範本總結：(一) 以上海市靜安寺及其周邊為尺度的研究範圍，應用密度分群法，根據取車點和還車點的分佈建立各自獨立的獎勵機制分區，計算最終每一輛單車的獎勵額度，促使使用者參與到調度的過程中。(二) 而以上海四環(121.75°E-121.125°E, 30.95°N-31.45°N) 為尺度的研究範圍，應用 DBSCAN 分群法，根據還車點的分佈建立密度視覺化監督系統，使得運營決策者能夠及時瞭解單車在不同密度值下呈現的分佈狀態，並以此做參考，來決定下一階段新投入運營單車的數量和投放地點。

第五章 討論與建議

5.1 調度策略總結

5.2 後續研究建議

第五章 討論與建議

5.1 調度策略總結

實作發現，在僅考量不同時間段內單車的流量變化時，若以上海市靜安寺周邊為研究範圍的情況下，則採用密度分群法作為獎勵分區得到的單車流動趨向較為合理，會使實施後的結果趨向于平衡；若以四環（ 121.75°E - 121.125°E ， 30.95°N - 31.45°N ）為研究範圍的情況下，則採用 DBSCAN 分群法作為決策者掌控單車密度分佈的方式，及時為調度決策提供實時資料支援。實際運用時應結合所面對的現象的具體情況，根據不同分群方法的優缺點，採用合理的分群方法處理資料。

5.2 後續研究建議

本研究所採用的上海市摩拜單車資料精確度不高；且2017年 共用單車與城市發展白皮書中的單車數量可能和實際運營的單車數量有出入。導致實作中出現許多無法藉由機器學習聚類分析解決的問題，尚需通過對方法論的探討、深究，運用策略（博弈論）的思維得出結論。

未來，隨著資料透明化的推行，精確度更高的資料可以公開使用到學術領域，大量資料的導入訓練，可增強本研究結果的分區準確度和更短週期的獎勵方案，使獎勵機制、調度策略更加具體。

對於基於閾值控制思維的逆向疏散獎勵機制，未來可將獎勵額度的分配做更細緻化的調整，例如引入餘額遞減法（Declining Balance Method）的思維分配每一個時段內的獎勵額。

對於在實際應用中探究三種聚類分析方法及其合適使用情境，未來可探討更多屬於機器學習領域的演算法，比如引入決策樹（Decision Tree）思維、異常檢測（anomaly detection）等構建單車流量控制系統，通過比較得出更適合某一個場景的演算法。

同時，在考量數據資料所呈現的單車狀態以及城市交通狀態時，可適當加入部分人文與客觀的因素，在地化考量上海市交通的實際狀況，比如單車道的普及率、居民選擇單車出行的意願等等，以增強案例的實際性和可探討性。

參考文獻

中文

王甯、杜豫川,「社區居民適宜步行距離閾值研究」,交通運輸研究第一卷第二期, pp. 24-27, 2015。

北京清華同衡規劃設計研究院,「2017年 共用單車與城市發展白皮書」, pp.3-5, 2017。

朱瑋、何京洋、王德,同濟大學城市規劃系,「法國公共自行車系統佈局方法與實證研究——以巴黎和里昂為例」, pp. 1-2, 2015。

雷小鋒、謝昆青、林帆、夏征義,「一種基於 K-Means 局部最優性的高效聚類演算法」,軟體學報第十九卷第七期, pp. 1683-1692, 2008。

英文

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise”, *International Conference on Knowledge Discovery and Data Mining*, pp. 226-231, 1996.

Fraley, C.; A. E. Raftery, “Model-Based Clustering, Discriminant Analysis, and Density Estimation”, *Journal of the American Statistical Association*, Volume 97, Issue 458, <https://doi.org/10.1198/016214502760047131>, 2002.

Lander, J.P., *R for Everyone: Advanced Analytics and Graphics*, Addison-Wesley Professional, 1 edition, December 29, 2013.

機器學習應用於共用單車流量分佈

Li, Y.X., Y. Zheng, H.C. Zhang, L. Chen., “Traffic Prediction In A Bike-sharing System”, *23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems Seattle*, No.33, pp.78-88, 2015.

Micheal, D. E., Mike W., “Bayesian Density Estimation and Inference Using Mistures”, *Journal of the American Statistical Association*, Volume 90, pp. 577-588, 1995.

Richard J. Roiger, Michael Geatz, *Data Mining: A Tutorial-Based Primer*, Addison-Wesley, 2002.

Speck, J., *Walkable City*, pp.3-6, 2012.

Timothy, H.,Zandbergen P., “Kernel density estimation and hotspot mapping: Examining the influence of interpolation method, grid cell size, and bandwidth on crime forecasting”, *An International Journal of Police Strategies & Management*, Volume 37, Issue 2, pp.305-323, <https://doi.org/10.1108/PIJPSM-04-2013-0039>, 2014.

網頁

CitiBike System Data, [cited 15 May 2018]. Available online:

<https://www.citibikenyc.com/system-data>

DBSCAN, [cited 15 May 2018]. Available online:

<https://zh.wikipedia.org/zh-tw/DBSCAN>

Mobike Global, [cited 15 May 2018]. Available online: <https://mobike.com/global/>

NYC.opendata, [cited 15 May 2018]. Available online:

機器學習應用於共用單車流量分佈

<https://data.cityofnewyork.us/view/ba8s-jw6u>

Ofo Riders, [cited 15 May 2018]. Available online: <https://www.ofo.com/us/en>

Soda Open data, [cited 15 May 2018]. Available online:

<http://shanghai.sodachallenges.com/data.html>

TechOrange, [cited 15 May 2018]. Available online:

<https://buzzorange.com/techorange/2017/08/21/share-everything-together/>