

摘要

YouBike 在台灣目前是公共自行車服務業者中最成功的，近年來在全台灣對於租借自行車的接受度也越來越高，但也因為大眾的需求增加，許多站點常常在尖峰時刻會有無車租，或是無車位還的問題，導致客戶滿意度下降。捷安特公司為此也相當煩惱，因為每天尖峰時間會有上述問題的站點都雷同，但都無法很準確的預估要配送多少的數量。針對此問題，最直覺的方式是在每個站點都建立一個模型來進行精準預測，但我們必須了解到，這樣的作法卻會為負責預測的單位帶來龐大的運算負擔，因為像光臺北市就約有 400 個租借站，若我們要在每個運算週期同時對這 400 站做預測，則其硬體需求將是非常大的，有鑑於此，本研究選擇運用集群分析找出租借量差異最大的前幾名站點來建立模型，再利用隨機森林判斷其他站點與主要站點哪個最相近，最後我們會計算其他站點與主要站點的預測誤差，並利用這些誤差建立低成本的外掛模型。如此我們在預測其他站點時，便能利用主要模型加上外掛模型來完成這些站點的預測，而這樣的預測成本預期會遠低於所有站點都建立一個完整的模型，並因此讓即時且低成本的預測所有 YouBike 租借量這件事變得易於實行，最終我們會實際與臺北市 YouBike 資料來驗證此方法的有效性。

關鍵字：YouBike、集群分析、隨機森林、外掛式模型