

# 大數據

# 縱橫路左右駛出來

## 計運輸空間時限 「最優分配」郵遞員

數據的運用與處理在不少人心中，只是技術而非科學；但面對極龐大的數據，程鴻(見右圖)肯定地說，這絕對是科學，並反映在如何處理這些數據上，「這包括了存儲、管理、計算等，而且要最快地找到結論，夠不夠高效、能否找到一個最好的算法，讓複雜的問題幾秒搞定，這個就是很科學的事情。」

到底數據有多龐大？程鴻嘗試口頭地把一個城市數據化，「大城市裡面的路網，你可把它看成公路網、路段、一個個地點、建築、遊覽點等，還有邊線和道路。這些都是非常豐富的信息。當中還有流動的資訊，像行人、交通等，加起來就是非常龐大的數據。」



大城市公路網蘊藏非常龐大的數據。設計圖片

### 物流研究非平面 須考慮「時地人」

她和微軟公司合作的研究，以北京四環到五環的交通地圖為藍本，製成標繪圖(Graph)，將它變成上述的數據化資訊——路段、點、建築、遊覽點等，還有邊線和道路。她進一步解釋說：「我們的項目，就是針對物流問題來進行，研究派遞公司如何提升效率。舉例說，每家公司有很多個郵遞員，每個郵遞員可能平均都要派送20個包裹，其間還會一直有客戶請求，叫派遞公司去上門取件，而且這些請求還是有限時的。面對這麼複雜的狀況，我們的研究就是希望用智能方法去決定安排哪一個派遞員去哪裡取件。」

程鴻指，傳統而言，派遞公司就是簡單以地區劃分，「但這不一定是最佳的做法，因為派遞員的任務數量不一定一樣，所以要用計算上比較高效的方法去做。」

她所進行的研究，並非外界所想像般只在一個平面上，將點和點連成直線，還要顧慮到很多元素。她解釋說：「派遞員身處於不同的地區，而且時時刻刻都在移動，而客戶的請求則如水流一般不斷湧入，哪一個甚麼時間在哪裡，算得快之餘，還要分配優，這樣才能降低成本，也同時能處理更多請求。本質是挺難的，連在路線上的哪一段去彎路也要考慮到。」

此外，物流業真正體現大數據的數據產生又快又多的特色，其郵遞員數量多，請求產生速度、數量亦多，「不可能每一個包裹劃一算20分鐘，所以要有效率高的技術。」

### 先過濾「不可能」提算速

程鴻的做法是，由一般公司的經驗分配，改為交由電腦科學去處理問題，「怎樣才能算得快呢？若沙田的取件請求，總不能叫中環的郵遞員去接吧？所以第一步就要做過濾。我們透過計算「空間時限指數」，就可以很快地過濾掉不可能的派遞員，從而增加計算速度。」她預計，有關做法能令一個複雜的分配問題於幾秒內產生答案，「其實這些都是經典問題，只是大家都想用計算方法，讓問題更好地解決。」

她表示，有關研究的算法已設計了，也取得北京公路網的數據，再用假設去產生人工數據去做試驗，「之後還要做測試，看速度夠不夠快。而且系統要能夠處理不同工作量，不論是1,000、1萬或10萬件工作。」此外，她更希望稍後找物流公司去交流，看看能否真正處理現實需求，有沒有其他限制或困難。



梁美兒希望透過研究電車行車數據推算其遇上特別情況的即時回饋方案。香港文匯報記者劉國權攝

## 令電車更「動」靈活應對突發

### 改善交通

大數據其中一個有趣之處是幾乎與任何事物都能扯上關係，例如大家認為變動性不大、走在軌路上的電車。香港中文大學系統工程與工程管理學系教授梁美兒與電車公司合作，希望透過研究電車的行車數據，以及透過手機應用程式收集乘客的需求數據，推算出電車遇上特別情況的即時回饋方案，作出即時調動，改善交通規劃。

梁美兒表示，交通規劃是很多地方都面對的問題，例如交通擠迫的歐美有很多行人區，當中應如何協調都是社會積極尋求答案的議題。放在香港這個人多車多、環境狹小的地方，更是挑戰重重，「香港行雙軌，又有穿插的路段(cut across)，可能會複雜一點。」

而她的研究目標就是更好地協調電車與乘客之間的編排，以應對各種突發情況，包括塞車、封路等，或賽馬、遊行等可預見的情況，以及做好日常交通安排，「將來每人都有智能手機，是否可以自動感應，以掌握整體市區中心的交通狀況，知道每個人想從哪裡去哪裡呢？如果可預先知道每人的目的地，對車輛調動、線路的編排和朝向等，都可有調節。」

### 盼製實時回饋系統

有關研究只是剛剛開始，目前梁美兒已從電車公司取得供應方的數據，「路線的分布、時間等這些資料都有。現時電車公司的做法是，6個月前設計了路線，今日開車，遇上狀況時再即時調配車輛。我希望可以在繁忙及非繁忙時段去設計

一個行車調配模型，並造出一個實時的回饋系統。」

不少人認為，電車的變動性不高，梁美兒亦表示，從表面看，電車可調配的東西的確不多，「因為電車涉及車路的限制，但也因為如此，挑戰性很高。但其實電車有很多分段的地方，讓我們可以決定甚麼時候要盡快調頭，可以得知下一輛車何時到，亦可以預算它去到哪裡。我們希望了解即時反應調動可以怎樣做。」

從微觀角度看，當中的變數又遠超我們想像，梁美兒解釋：「司機的下班時間、午膳時間等，多餘的車輛又應該放在哪裡，這些都需要一一部署。」

### 最難獲取乘客去向

但最困難的是如何取得乘客的去向資訊，梁美兒坦言：「現時未有這樣的技術，但乘客的分布對如何安排路線十分重要。」現時研究團隊在討論不同方案，「可能會出個App(手機應用程式)」，現時也有應用程式說下一輛車何時會到。我們也想過用電話公司的網絡，又或者在車站加裝攝錄機，讓我們可以知道多不多人在等車，不過這樣的收取與分析不容易，當中也涉及私隱問題，攝錄機若要定期維修，也要成本。」

在研究出最終對策前，梁美兒想出一個折衷方案，就是以往站與站之間的數據去推算，了解擠迫時可做些甚麼，再加以分析，「希望將來可取得需求方的數據，用香港實況去寫出模型。」

歐陽文情

香港文匯報訊(記者 歐陽文情)大數據無處不在，除上一集提到可應用於教育外，對於縱橫交錯的物流路線規劃亦大有可為。香港中文大學系統工程與工程管理學系助理教授程鴻與微軟公司合作，透過分析路面數據及不同的流動數據，計算物流路線，讓四出派遞並接收包裹的郵遞員可在工作路線上作出最合理分配。研究團隊更運用「空間時限指數」(Spatial Temporal Data Base Index)去作出快速篩選，提升運算速度，過濾出「最優分配」。



### 萬物互聯

助選最快個人路線  
上街省時快捷

或許大家對北京的物流公司路線編排不感興趣，但程鴻說，有關算法只要研發成功，其實可套用於任何一個城市。此外，這種路線編排除擴大至物流公司外，亦可縮小至個人層面，為大家日常出行編排出最優路線，甚至發揮其他功能，透過關鍵字極速搜尋身邊各類的「Top 5(首五位)」，將大數據「萬物互聯」的本質發揮到極致。

大家每日出門或有好幾項任務，須去數個不同地方。程鴻表示，其研究有助進行「個人路線搜索」，「你首先要找出哪幾個point of interest(關注地點)，比如說我今天出行，從家裡出發，要去到銀行、超市、郵局，除了家以外，其他地方的網點不是唯一的，可以去不同地方的銀行、超市、郵局。如果我決定去沙田某商場的超市，但銀行未必在附近，我可能最終要走更多的路。所以如何安排路線，其中一個參考，就是看這幾個不同點的「聚集程度」，並且用關鍵字眼去做個人化的搜尋。」

### 空污減碳都得 環保「新星」

同時，有關研究亦可發展其他應用層面，程鴻說：「比如說我想找5所離我家最近、而且是英文教學的小學，如何算得快，如果從廣度優先搜，就會慢，反而應該建立計算的指數，再漸漸擴大。」這種交通路線編排對程鴻而言，不單節省交通時間或走少一點路，還可幫助控制空氣污染，減低碳排放，所以在環境保護上亦有幫助。

## 拼車App 即搵最近的士

### 初師高手

交通或許是最早能在商業層面應用大數據的領域之一，現時內地的士相關的手機應用程式盛行，除有幫忙「打的」的應用，還有幫忙「拼車」的應用。程鴻表示，其中微軟開發一個「拼車」打的軟件，「高峰期打車的人很多，車不夠，於是就要拼車，而那個軟件就是幫忙找到最近的的士。」

香港科技大學霍英東研究院院長倪明選亦表示，在交通方面，上海、北京、深圳都收集齊全的士數據，「收集得非常仔細，可以知道幾時幾分，的士在甚麼地方、有沒有載人、速度多快，香港在這方面已落伍。內地是政府要求做的，有了這樣的數據，可以做很多事，比如說看它的速度變化，算出來甚麼時候哪裡塞車最厲害。」此外，透過的士的行車路線可知道甚麼時候哪條路線最好，該段路線需多少駕駛時間，都可分析出來，「地圖算路線只是找最短，有沒有塞車也不知道。」