

# 分析大數據 預測防罪案

— 今日香港 + 全球化 —

現時經常聽到人講大數據，由於海量資料中隱藏着許多高價值的資訊，比別人愈早發現，其價值就愈高。在現實中，大數據早已普遍運用在商業、管理和科技研發等領域上。當中以在商業上的應用最廣為人知，專家對大數據作出分析，建立數據模型進行預測，找出顧客消費的習慣和行為，例如準確預測流失客戶的特性和時間，從而調整公司的產品研發或銷售策略。大數據的確為商業、管理和科技研發等領域上帶來很多前瞻性的預測，但很多人並不知道，很多警察機關，例如洛杉磯、東京、柏林、芝加哥、阿姆斯特丹、倫敦和紐約都已利用大數據來分析罪案的發生，把大數據應用到預防和打擊罪案的範疇。

■余慕帆 前線通識科老師



大數據可以應用在不同方面，圖為2017中國國際大數據產業博覽會。資料圖片

## 查犯罪記錄 測時間地點

現代科技日新月異，網絡科技犯罪的案件經常出現，導致網絡安全及科技罪案等問題一直上升。各國的警察部門正借助科技的力量，處理有關網絡安全的問題及調查科技罪案，包括處理電腦法理鑑證及防止科技罪案。但除了利用科技技術防範科技罪案外，在防止實體罪案方面，警察機關亦嘗試利用大數據的技術，以收集和分析海量大數據去推算罪案發生的時間、地點、罪案的類型、犯案者是誰等等的重要資料。讓我們先了解一下不同的警察機關是如何收集和分析大數據來防止或打擊罪案的發生：

- 1. 德國柏林**  
警方研發一個名字叫Precobs的軟件系統，這個軟件系統根據收集回來的不同的海量數據，包括過去的罪案數據和犯罪細節，作出分析，並嘗試從數據中找到關聯的模式，預測發生罪案頻率最高的時間和地點。德國警方把這軟件系統應用在入屋爆竊的案件分析和預測上，希望將來擴大到其他的罪案範疇。
- 2. 日本東京**  
日本以科技著名，現在亦嘗試利用每一輛東京的車子作為數據，從數據推算最多車禍發生的時間和地點，配以適當的措施，減少因車禍而導致死亡的案件數量。
- 3. 中國深圳**  
根據深圳交警的網頁顯示，深圳交警已將互聯網思維、大資料技術與交通管理業務相結合為4.0新模式，以大資料研判為基點，根據過去各項交通違法數據進行分類歸納和分析，從中發現相關性，找出最多違法車輛出現的路段，並根據預測結果加派人手於該路段進行檢控。檢控數據又會回饋到大數據中，使將來的數據分析更加準確。

## 概念鏈接 什麼是大數據？

大數據有三大特徵。第一是海量，資料量的單位可從一兆位元組到千兆位元組，是以前的科技和技術所不能處理的。

第二是多樣，大數據的來源可以很多，從社交平台的文字、無線網絡的連接狀態、到監控鏡頭的影像都可以是數據來源。

第三是高速，大數據的資料不斷即時更新，更新的數據會高速被儲存和處理。

- 4. 中國上海**  
上海市公安局與當地電信營運商合作，取得和收集人流密集區域內的移動電話通訊基站數據和無線網絡的連接記錄，從數據中分析人流量，再結合現場錄像監控以識別結果，向公眾發佈人群聚集風險的預警提示。
- 5. 美國洛杉磯**  
洛杉磯警方與大數據分析服務公司Palantir合作，研發了一個搜索挖掘平台。警方可通過搜尋輸入罪案、拘捕記錄或文件等詞語進行搜尋。這個平台把多個數據庫的數據連接，連久遠的數據也能搜尋到，處理海量數據的速度驚人，有效為洛杉磯警方尋找有關犯罪的資訊。
- 6. 美國加州**  
美國警局透過收集過去的警方資料，並運用數據分析，研發預測性警務軟體PredPol，目的是預測有可能發生罪案的機率。研究團隊根據三項因素推算出未來案件出現的機率。第一，犯罪案件經常出現的地點如夜店。第二，多次受害地區如屢次遭人爆竊的地區。第三，受害地區的鄰近地區會有較大機會出現類似的案件。鎖定未來較大機會出現案件的地點後，警局會對執勤的安排作出調整，便能預防案件的發生。

## 減少犯罪率 巡邏更有效

- 1. 有效遏止罪案發生，減低犯罪率**  
以美國為例子，在加州內，採用PredPol軟件系統的地區，入屋爆竊的案件比其他地區減少了大約11%，搶劫案件亦大幅下降了27%。美國洛杉磯在使用PredPol後，罪案率亦降低了13%。
- 2. 更有效運用資源**  
美國洛杉磯警方在使用預測性警務軟體後，能借分析數據，有效地在某些預測將有罪案發生的地方加派警員巡邏，阻止罪案發生或第一時間到達案發現場。
- 3. 減少傳統人力資源的運用和時間**  
深圳交警透過大數據分析，配合鐵騎隊的巡邏，有效發現在主要幹道的違法者，亦縮短了傳統人力資源在主要幹道的時間。瑞典公司QlikView研發用作情報分析的工具平台，此平台
- 4. 增加數據的切入面，快速發現數據的關聯或模式**  
不少數據系統擁有可視化介面，可以把多個數據庫的數據進行連結，讓人可以由任何一個數據庫作為切入點，與其他數據進行合併分析，增加預測的準確度。



在南京舉行的軟件產品和信息服务交易博覽會上，參觀者在觀看大數據信息應用展示。資料圖片

## 憂侵犯隱私 易偏頗標籤

- 1. 成本高**  
利用大數據減罪需要收集海量的資料，從數據中發現特有的模式或找出關聯和進行預測。大數據的收集並不容易，亦需相關領域的專家針對數據進行分析和預測。再者，研發軟件系統的費用高昂，此後亦可能每年需要繳交昂貴的服務費。以美國的PredPol為例，每年每個警局都需要交大約十萬美元或以上的服務費。
- 2. 侵犯隱私**  
有地方政府與電信營運商合作，取得和收集人流密集區域內的移動電話通訊基站數據和無線網絡的連接記錄，並在重點區域出入口現場錄像監控，統計數據，分析出監控區域內的人流量。有人憂慮個人的資料如個人的通訊資料和無線網絡的連接記錄在未經同意下外流，個人隱私得不到保障。另外，監控鏡頭技術日新月異，當中人臉識別被用作計算民眾活動及行為，例如售賣武器店的監控鏡頭能助警方調查可疑人物，但有人卻怕市民大眾可能在不知情的情況下被收集肖像。
- 3. 標籤效應**  
有指美國警方根據數據分析得出高危險名單，有不少被捕者早已在名單上，但有人質疑由軟件系統從數據預測的準確性，質疑證明軟件系統有效預測罪案的研究不足，害怕警方在收集數據時已經出現偏頗的情況，讓數據本身代表性不足，影響分析結果和準確性。例如有人認為部分黑人和貧民聚居的社區特別受警方監察，讓當地民眾覺得受標籤和針對。



各地都在發展大數據產業，圖為內蒙古大數據產業的設備。資料圖片



深圳交警使用大數據分析打擊罪案，圖為深圳市交警局福田交警大隊紅樹林扣車場。資料圖片

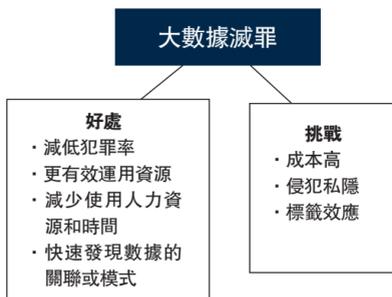
### 想一想

1. 什麼是大數據？
2. 利用大數據防止罪案發生有什麼好處和壞處？
3. 大數據的特徵是什麼？
4. 你認為應用大數據能有效減罪嗎？

### 答題指引

1. 本題問大數據的定義，同學可指出大數據就是大量的資料，透過分析這些資料，可以得到不同的資訊。
2. 本題可以根據資料作答，指出大數據防止罪案發生方面，可以減低罪案率、減少人力資源使用等優勢，但同時壞處包括成本高、侵犯隱私等問題。
3. 本題亦可根據資料回答，大數據主要有三個特徵，即海量、多樣和高速。
4. 本題屬於開放評論題，同學可自由作答，如同意的話，同學可指出大數據可以令執法機關找出罪案的時間和分佈，以及在犯罪前發現痕跡，令警方可以在罪案發生前有效預防。若不同意的話，可指出大數據的使用有其限制，由於數據量太大，執法機關未必可以在罪案發生前發現痕跡，即便發現，也可能被忽略，導致未能阻止罪案發生。

### 概念圖



### 延伸閱讀

1. 《馬培華籲適時為大數據立法》，香港《文匯報》<http://paper.wenweipo.com/2017/10/19/CH17101900060.htm>
2. 《eBay靠大數據推測大熱產品》，香港《文匯報》<http://paper.wenweipo.com/2017/10/13/FI17101300117.htm>
3. 《大數據平台計劃兩年出爐》，香港《文匯報》<http://paper.wenweipo.com/2017/10/19/HK1710190036.htm>