

# 大數據

# 輪轉市上下計出來

## 找「高危」保險客 推導最佳理賠方案

理工大學互聯網及移動計算實驗室主任曹建農表示，從保險公司每年的顧客賠償個案，「可以看到每個人花費不一樣，有人住院3天，有的要住10天。一般這些保險公司會每年作分析，當中哪些人賠最多錢，以後就要加保費；還有工傷的治療方法建議，哪一種工傷、哪一個治療康復更快、用錢更少等。」

曹建農其中一個研究，就是替30多家保險公司以大數據去分析保險賠償的申請原因、賠償方法、治療過程等資料，透視當中現象並找出對策，「雖然現在只有幾萬個數據，可是如累積到過去5年至10年的數據，可以看到賠償及治療技術的變化。」研究開始逾半年，其中一個目標是評估「風險指數」，找出高風險的人，「危險度跟性別和工種有關，比如說男的、在室外工作的，受傷機率會高一點；女的、在室內的，精神上可能有問題的機率多一點。」此外，項目還會透過數據總結出「最好治療方案」，指定他們用特定的治療方法才獲賠。

### 「考古」看紀錄 查受保人有否欺詐

科技大學霍英東研究院院長倪明選及研究助理教授譚浩宇亦分析客戶行為，課題亦是保險業。他們與內地一保險公司合作，「我們希望找出每一次客戶索賠時，有沒有欺詐可能。保險公司不可能把每一宗都查出來，但假設每天有1,000起案件，如果我只是查100起，怎樣才能找出最大可能的欺詐？」

他們就把這些個案排序，先查榜上靠前的，二人解釋說：「這是根據客戶過去的行為做排序。傳統做法是用最簡單的信息包括案件、發生地點、車型等。現在則加入人的『歷史行為數據』，包括投保歷史、開車歷史、道路本身事故發生頻率、天氣、環境等，都可放到模型去提高準確度。」

譚浩宇又直言，單一數據模型並不足夠，要加上不同模型，才可更準確分析，「現時我們的計算比美國知名諮詢公司要好一點，但我們計算資源的成本更低，對中小企業而言很有競爭力。」

### 金融業也適用 助定信用評級

除了保險，倪明選和譚浩宇均表示，大數據對互聯網金融亦有很大用處。譚浩宇以「小型貸款」為例說：「阿里巴巴做了『阿里小貸』的產品，把幾塊到幾萬塊的貸款發放給旗下『天貓』（購物網）的店主。為何銀行不做這些？因為要查資料、填表等，成本比例太高。阿里巴巴沒這問題，因為它已有店主和店舖的行為數據，可算出信用評級；而且它透過網路就可，軟件也是開源軟件稍作修改就可。結果他們把錢發放給自己的客戶，能做到一塊錢貸款也可賺錢。」

他們認為有關模式亦可在港開創商機，「香港有不少小型財務公司，一般而言都要客戶的收入、報稅證明，再去算借貸額，過程繁瑣。若有一個機構，有我的交易歷史紀錄，變成信用評分產品，這些公司透過網上查詢就可決定是否貸款給我。內地就有很多這樣的事情。這些數據在大銀行裡，若能說服大銀行把數據拿出來，做成產品，服務第三方的小型金融企業，這樣數據就直接變成資產。」



曹建農研究的其中一個目標是評估風險指數。  
香港文匯報記者莫雲芝攝

## 開發工具 最難是錢

### 普及大眾

有科學家研究大數據不同的應用範疇，也有人開發供一般採用的大數據工具。理工大學電子計算學系副教授盧至力，就以開發大數據分析工具為目的，讓不太懂得電腦的人都能用到，就像現時各式各樣的電腦軟件，人們即使不清楚背後原理，亦可有效運用並達到指定目的。

### 運用問題：數據既急又多

盧至力概括大數據運用時所面對的幾個問題，「就是數據很多、來得很急、有很多不同類型的數據，例如我們買東西，會有一行行的購買紀錄，包括何時及買了甚麼；而現在買完東西又會去網上社交平台說產品好不好，這不是單純一行行資料，而是一堆描述文字。但不同人的說話又會有不同分量，我和行政長官梁振英說的就完全不同，因為他有更多人關注，這點不是數據紀錄，也不是自由文字，而是人和人之間的網路圖，這就有3種資訊要看。傳統公司不懂得如何分析，我就做個系統，將麻煩東西收集起，讓他們更方便地進行分析。」

他希望開發出數據工具，不同用家寫幾句東西，無論最終數據有多大，都可給出結果，「如果數據很多，我們可於伺服器多加100台機或1,000台機，替它自動化地做分析。」

但他坦言，大學做有關工具開發面對不少難題，「最困難是錢，此外我們的規模亦遠未及外面的大公司；大數據講的是幾百萬台機，大學做不到。所以我們傾向與企業合作，因為他們有硬件資源，但未必有人和技術，我們則相反。」

歐陽文情

## 縮短研藥時間 福澤病友

### 數碼仁心

大數據還能應用在醫療方面，理工大學電子計算學系教授陳振沖將大數據用於藥物設計，透過電腦運算去應付各式各樣的配對工作，以縮短研製藥物的時間。

### 內地出動「超腦」配對

陳振沖解釋說：「當人有病時，可能是某個蛋白有缺憾，藥物就針對該蛋白去解決問題。但人體的蛋白種類很多，要先看有甚麼病、會有甚麼影響，而藥物也可能會有副作用等，當中有很多可能性，有很多數據要分析。內地也會用超級電腦來做。」

他坦言，現時研發一種藥往往要十多年，當中很多

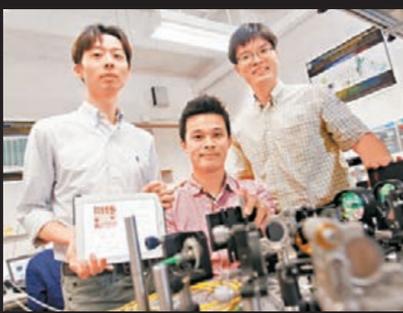
時間用來做實驗，但若先用電腦做分析數據，就可快點配對出來。研究團隊首先了解指定蛋白有何特徵、合適藥物又有甚麼特徵，「每隻藥有幾十個至幾百個副作用，而在不同人身上也可能略有不同，現時數據很多，只是沒有方法分析；我們最終想瞭解為何有些藥會有這樣的副作用，及到底甚麼藥應用在甚麼人身上。」

雖然研究暫時未有成功例子，但他相信鑽研下去最終會令研製藥物的周期縮短，「現時我們看到一些東西，但要有生物學意義的，還需探究，並要找有藥物開發經驗的人參與。現時很多藥物資料數據都已公開，能有更多當然更好，而涉及病人的資料則要特別收集，可能要與藥廠合作。」

歐陽文情



陳振沖希望透過大數據的電腦運算配對來縮短藥研時間。  
記者莫雲芝攝



港大電機電子工程系團隊探討以FPGA結合ATOM，令處理高清細胞圖像提速最高10倍。  
記者劉國權攝

## FPGA X ATOM 更快「捉」癌「細」胞

### 高清图像

當社會對資訊的要求越見精細，處理數據量也更龐大，電子儀器運算速度成為其中一個技術關鍵。香港大學電機電子工程系團隊，正探討以高能的「現場可程式邏輯門陣列（FPGA）」代替傳統電腦，結合學系研發的「ATOM」光學高速顯微成像系統，令處理高清图像的速度提高5倍至10倍，讓技術更切合醫學如癌細胞病變診斷需要，造福大眾。

港大「ATOM」系統將光纖光學和激光技術結合攝影概念，成功令小至細胞層面的圖像拍

攝速度及精細度大幅提高，即使在人體組織只有極少量癌細胞，也可成功捕捉；但由於所拍得的高清圖像涉及超大量數據，如何有效處理及篩選圖像，避免「拍得太快跟不上」，成為系統運作的重大挑戰。

負責處理FPGA技術的電機電子工程系助理教授蘇國希介紹說，傳統電腦的設計在於多種形式的全面運算，而FPGA則是已改變構造的電子硬件，透過特有的程式令所有運算資源「專門化」，從而提高速度及效率。而有關做法對個別種類大數據的處理也有較大應用潛力。

香港文匯報記者 任智鵬