

本期焦點 | 所長的話 2 | 名人榜 3 | 深度報導 4 | 卓越講座系列 6 | 活動快訊 7 | 計畫特寫 8 |  
實驗室報導 10 | 創意發想 12 | 人物焦點 14



# IIS

資訊科學簡訊



# Update

Newsletter of the Institute of Information Science, Academia Sinica, Taiwan

2017 年

資訊科學研究所於 1977 年開始設立籌備處，歷經五年籌備，於 1982 年 9 月正式成立研究所，是中央研究院數理組十個單位之一。本所的研究重點分為八大方向，包含生物資訊、電腦系統、資料處理與探勘、多媒體技術、語言與知識處理、網路系統與服務、程式語言與形式方法、計算理論與演算法。

本所並於 2003 年配合中研院國際研究生院 (TIGP)，開始招收生物資訊學程博士班，致力延攬及培訓此一新領域之國際人材多年有成，已招募 65 餘名學生，對生物資訊的研究有很大的助益。2014 年起，社群網路與人智計算新學程博士班開始招生，目前已招募 25 餘名學生。

所內資深研究人員皆指導或共同指導來自全國各地頂尖大學之碩、博士生，並與這些大學保持緊密的研究合作關係，與國內外資訊業界亦有極為良好之互動關係。

所長：許聞廉

副所長：王新民  
劉庭祿

研究群主持人：

施純傑教授

生物資訊實驗室

吳真貞教授

電腦系統實驗室

陳孟彰教授

資料處理與探勘實驗室

廖弘源教授

多媒體技術實驗室

許聞廉教授

語言與知識處理實驗室

陳伶志教授

網路系統與服務實驗室

穆信成教授

程式語言與形式方法實驗室

王大為教授

計算理論與演算法實驗室



**首**先，歡迎新任助研究員蘇黎博士加入本所研究團隊，也恭喜楊得年博士升等為研究員。

時光飛逝，在此暑往寒來之際，上一期通訊發行已年餘。這年以來，本所同仁持續致力於資訊相關領域的研究工作，獲得豐碩的研究成果，並有多位同仁榮獲國內外眾多獎項與殊榮，包括前所長及特聘研究員李德財院士榮獲「伊利諾大學傑出校友獎」、張原豪副研究員榮獲中國電機工程學會 106 年度「傑出電機工程教授獎」、鐘楷閔副研究員榮獲傑出人才發展基金會第五屆「年輕學者創新獎」、廖弘源特聘研究員榮獲第二十三屆「東元獎」。

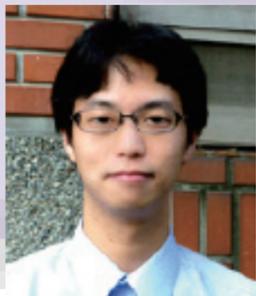
本期通訊的深度報導「社群媒體文字資料之人際互動網絡建構」，為本人領軍 IASL 團隊所進行之研究。在這個網路普及的時代，網際網路已經成為人們分享與取得知識的重要來源，且隨著社群多媒體的蓬勃發展，任何網際網路使用者都能夠自由地對任何事件發表評論與感想，這種便利性雖然讓網際網路成為了解天下事的重要知識寶庫，超載的資訊量卻也造成使用者的極大負擔，因此，本研究嘗試從主題事件文章中找出具代表性的部件暨其交互關聯性，讓使用者得以快速有效地綜觀事件的發展。

其次，計畫特寫闡述林仲彥副研究員所研究的「甲基化快速分析平台解析生物醫學」計畫，獲得科技部與微軟公司全球雲端研究計畫贊助，並首創原始定序數據圖像化呈現完整分析平台；再者，實驗室報導介紹程式語言與形式方法實驗室，由穆信成副研究員引介如何以邏輯與數學為工具，運用形式化的方式確保軟體之正確性；而創意發想的園地，廖純中研究員解析了一個關於多數決投票的謎題「法官們的困境」(Doctrinal Paradox)，藉由反省這個謎題，觀察到多數決的一個瑕疵；最後人物焦點部分，則介紹目前正致力於「空氣盒子專案」，開發以群眾參與為基礎的大型 PM2.5 感測系統的陳伶志副研究員。

希望能藉由本期通訊的內容與您交流，並請您一如往常提供寶貴的建議及回響，更竭誠歡迎您抽空蒞臨指教。



# 名人榜



楊得年博士升等為本所研究員，自 105 年 9 月 22 日生效。

李德財院士榮獲伊利諾大學傑出校友獎。



副研究員張原豪博士榮獲中國電機工程學會「106 年度傑出電機工程教授獎」；以及獲選本院 107 年度前瞻計畫。



副研究員鐘楷閔博士榮獲傑出人才發展基金會第五屆「年輕學者創新獎」。

廖弘源特聘研究員榮獲第二十三屆「東元獎」。



## 卓越講座系列



Machine Learning Revisited  
**Eugene Wong** — October 3, 2017

".... Machine learning using neural networks is the rage of computer science. It is estimated that a large proportion of all computing cycles are consumed by machine learning applications. Even with parallelization and special processors, training is time consuming and laborious. A large application could take days and weeks to complete the training process....."

Social Physics: Toward a Computational Understanding of Society  
**Alex Pentland** — September 12, 2017

"New "big data" analytics now provide a computational framework for dramatically better understanding of human societies. This new capability is revolutionizing both governments and industry, but if it is to be most effective we need new approaches to data analytics, new "AI" capabilities,....."



Syntax-Guided Synthesis  
**Rajeev Alur** — December 19, 2016

"The formulation of the syntax-guided synthesis problem (SyGuS) is aimed at standardizing the core computational problem common to these proposals in a logical framework. The input to the SyGuS problem consists of a background theory, a semantic correctness specification for the desired program given by a logical formula, and a syntactic set of candidate implementations given by a grammar....."

# 名人榜

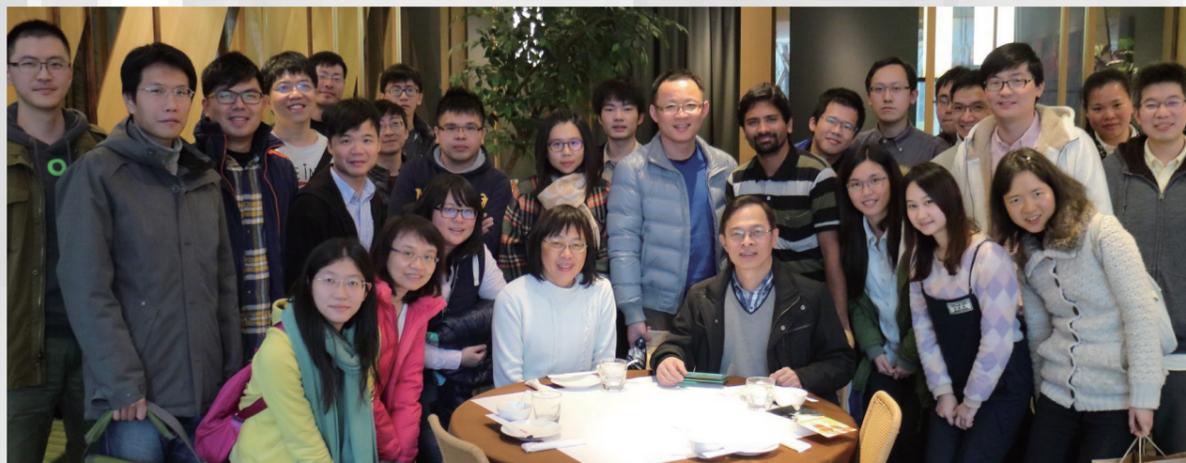


# 社群媒體文字資料之人際互動網絡建構

## 捕捉更豐富且精準的事件人物互動

張詠淳 / 陳建錦

許聞廉



Intelligent Agent Systems Lab 研究群。

在這個網路普及的時代，網際網路已經成為人們分享與取得知識的重要來源，且隨著社群多媒體的蓬勃發展，任何網際網路使用者都能夠自由地對任何事件發表評論與感想。這種便利性雖然讓網際網路成為了解天下事的重要知識寶庫，但超載的資訊量卻也造成使用者的負擔。為提供人們有效率的綜觀事件的發展，主題探勘 (topic mining) 近期成為一項重要的知識管理研究議題。通常，一個圍繞著某主題事件的文章，是由一個核心事件或行動、以及與之直接相關的其他事件或行動所構成。而一個主題 (topic) 則可歸納為原因、引發條件、特定時間、地點、參與的人或物等主要組成部分，另可能伴隨某些必然結果。因此，本研究之重點在於如何從主題事件文章中找出具代表性的部件暨其交互關聯性。此外，透過關係擷取技術，我們可以了解主題事件中人物的關係，進而可提供更深入的自然語言處

理與資訊檢索運用，故本研究也將分析主題事件中的人物互動關係。但由於人與人的互動關係會因不同事件而異，因此我們提出的方法為分析一事件的相關文件內容，且為每個事件建立專屬的社群互動網絡。然而，現行的關係擷取研究都僅就單一文句來擷取物件關係，但物件關係的描述往往可跨越數個文句。因此，本研究中採用人名於文件的位置作為關係文句片段的端點，以期分析每個片段內容並擷取適當的人物互動關係。我們預期此擷取方式能捕捉到更多且更精準的事件人物互動。

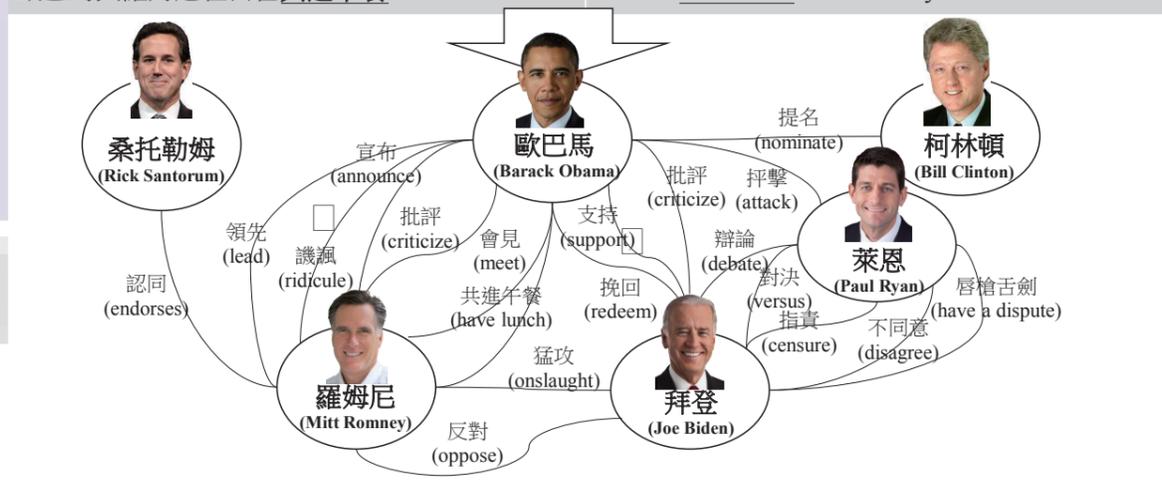
### 人物互動資訊的句型樹狀結構

本研究的方法基於一個富涵有人物互動資訊的句型樹狀結構 (rich interactive tree, RIT)，並透過此樹狀結構來捕捉主題人物互動關係。首先會先根據主題人物建構其最短路徑樹 (shortest path tree,

SPT), 然而, 根據我們的觀察, 經過 SPT 生成結果會將辨識主題人物 (如圖一中的歐巴馬與羅姆尼) 之間的互動關係重要資訊 (即「延攬」一詞) 給移除。針對此問題, 本方法將額外檢查剖析樹上的動詞片語結構, 進而決定是否要進行分枝 (branching) 的操作, 以找回被 SPT 刪除的資訊。如圖一所示, RIT branching 會延續 SPT 之結構以將遺失的重要資訊找回。此外, 為了讓 RIT 結構更為簡潔, 我們進一步移除 RIT 上對於辨識人物互動關係沒有幫助的的雜訊 (稱之為 RIT pruning), 進而獲得更為簡化之樹狀結構。此外, 本方法更進一步考量到互動關係中的動詞描述部分, 但並非所有的動詞皆可以用來描述人際互動關係, 故若能將互動關係詞彙加以標註, 勢必可以增加辨識之效能。有鑑於此, 本方法預期透過特徵選取 (feature selection) 技術從訓練語料集中建構互動關係詞彙清單, 對每個 RIT 結構中, 若含有互動關係動詞則會在根節點 (root) 增加額外的 IV tag 作為標註。

圖二為本研究方法所生成的 2012 年美國總統大選主題人物互動網絡。我們可從中觀察到, 民主黨候選人歐巴馬時常批評共和黨候選人羅姆尼的政見; 然

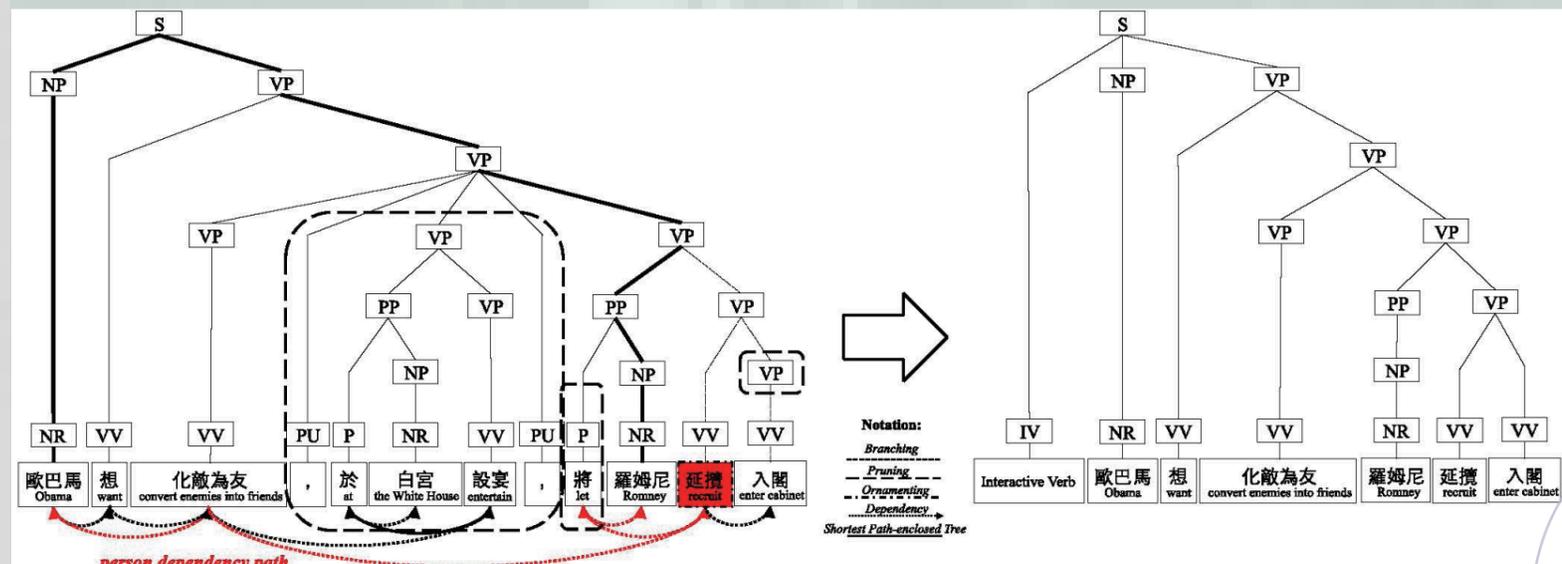
Chinese	English
美國總統歐巴馬以47%比46%領先共和黨總統候選人羅姆尼	U.S. President Obama <b>leading</b> Republican presidential candidate Romney with 47% to 46% of opinion polling.
拜登在首次副總統辯論上與萊恩唇槍舌劍	Biden and Ryan <b>have a dispute</b> in the first debate of vice president.
歐巴馬總統改變戰略, 猛轟對手羅姆尼	President Obama changes his strategies, and started to <b>criticize</b> his rival Romney
副總統拜登對決主打保守預算政策的共和黨副總統候選人萊恩	Vice President Joe Biden <b>versus</b> Republican vice presidential candidate Paul Ryan who promotes a conservative budget policy.
.....	.....
羅姆尼抵達白宮會面歐巴馬	Romney arrived at the White House to <b>meet</b> Obama
歐巴馬與羅姆尼在白宮共進午餐	Obama <b>has lunch</b> with Romney at the White House



2012 美國總統大選主題人物互動網絡 (圖二)。

而, 隨著大選結束後, 在歐巴馬組織內閣相關議題中, 他則是邀請羅姆尼一同共進早餐, 甚至邀請其擔任閣揆。由此可知, 他們的互動關係從負面 (批判) 轉為正面 (共進早餐), 也證實了主題人物間的互動關係是會隨著時間以及主題的不同而產生變異。因此, 本研究可發掘創新且具有洞見的成果。此外, 透過此視覺化的方式呈現主題人物互動

關係, 有助於瞭解主題資訊以及進一步探索人物間的競合關係, 這有別於現有的社群網路分析方法, 因為它除了提供正向的互動關係外, 更進一步提供了負面的互動關係, 以及用於描述該互動關係之關鍵詞彙。此結果除有助於讀者掌握主題的背景外, 更能提供另一種不同面向給使用者來瞭解此主題中之社群結構。



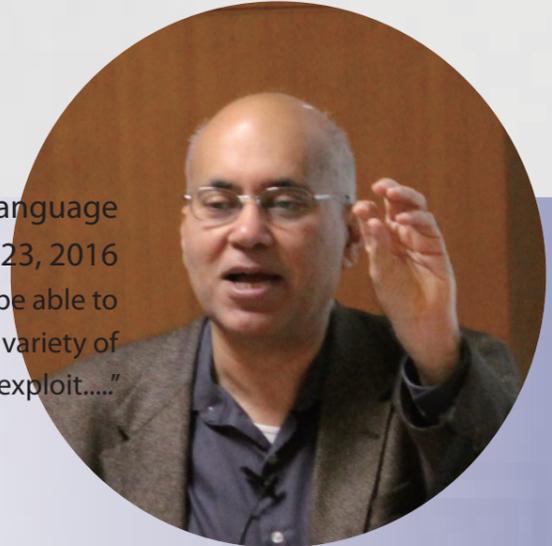
豐富互動關係句型樹狀結構生成過程 (圖一)。



Embodied Cognition: Towards Combining Vision, Motor Control and Language

**Jitendra Malik** — November 23, 2016

"We have been helped by "big annotation" and "big simulation;" but these will not be able to carry us all the way in other components of AI. Child development suggests a wide variety of learning mechanisms, which we can try to model and exploit....."



Revisiting Control/Data Plane Separation in Software Defined Networking

**Giuseppe Bianchi** — October 31, 2016

"This talk addresses technical ways to formally describe stateful per-flow behavior while retaining pragmatism (and some level of compatibility with today's OpenFlow), platform independency, and portability across different network devices and nodes....."

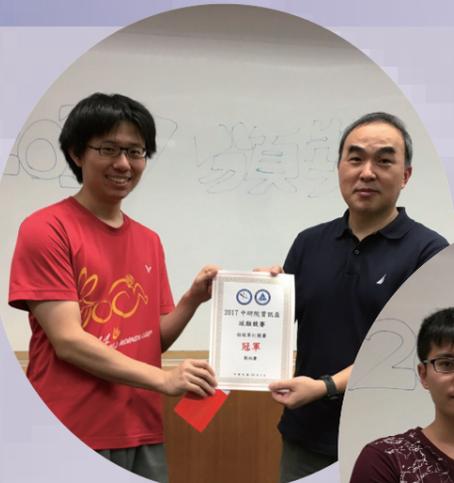
A Random Walk of Network Science Research:  
Sub-modularity, Large Graphs Computation and File System Support

**John C.S. Lui** — November 8, 2017

"..... how one can combine theory of sub-modularity and probabilistic counting data structures to information discovery and computation on large graphs. I will also discuss how one should re-design the file system support so as to provide a more efficient computation on network science problems....."



## 活動快訊



中研院資訊盃羽球單打冠軍得主劉柏慶



中研院資訊盃投籃王得主張詠淳



中研院資訊盃桌球單打冠軍得主侯人誠



資訊盃西瓜大胃王爭霸戰 45 秒完勝



院區開放活動闖關抽獎

## 2017 Frontiers of Communications and Networking Workshop

November 10, 2017  
Auditorium 106 at new IIS Building, Academia Sinica

To provide a platform for scholars in the field of communications and networking in Taiwan to interact, collaborate, and share new and trending ideas with each other annually.

### Topics

Realizing IoT & 5G Innovation and Services with SDN/NFV  
Sampling Large Networks: Algorithms and Applications  
Efficient Information and Influence Diffusion in Communications and Social Networks  
.....

## An International Seminar on Asia-Pacific Cooperative Security

May 26-29, 2017  
Auditorium 106 at new IIS Building

### Topics

- Security and Great Power Rivalry in the Asia-Pacific
- The Future of Missile Defense System
- Missile Defense
- Trump and the Security of the Asia-Pacific
- Aerospace Technology Development for Keeping the Peace Across the Taiwan Strait
- International Regulation of Cyber Sphere and Cyber Arms Control
- Challenging Myths of Cyber-Security and Cyber War--What Chance Cyber-Governance?
- Taiwan's Search for International Space in Regulatory Regimes
- Roundtable Discussion on Cyber Conflicts
- The Dilemma of the East and South (China) Seas
- Spent Nuclear Fuel Management
- Peace Museums
- Trust, Identity and Reconciliation in Northeast Asia: Dealing with Painful History to Create a Peaceful Present



## 2017 Frontiers of Communications and Networking Workshop

November 10, 2017  
Auditorium 106 at new IIS Building, Academia Sinica

To provide a platform for scholars in the field of communications and networking in Taiwan to interact, collaborate, and share new and trending ideas with each other annually.

### Topics

Realizing IoT & 5G Innovation and Services with SDN/NFV  
Sampling Large Networks: Algorithms and Applications  
Efficient Information and Influence Diffusion in Communications and Social Networks  
....



9月4日至10月20日  
共計33天的Tabata  
運動魂開始燃燒了!



6月16日年度自強活動於  
新竹馬武督白木屋渡假中心



10月28日院區開放活動資訊所吸引眾多參觀人潮



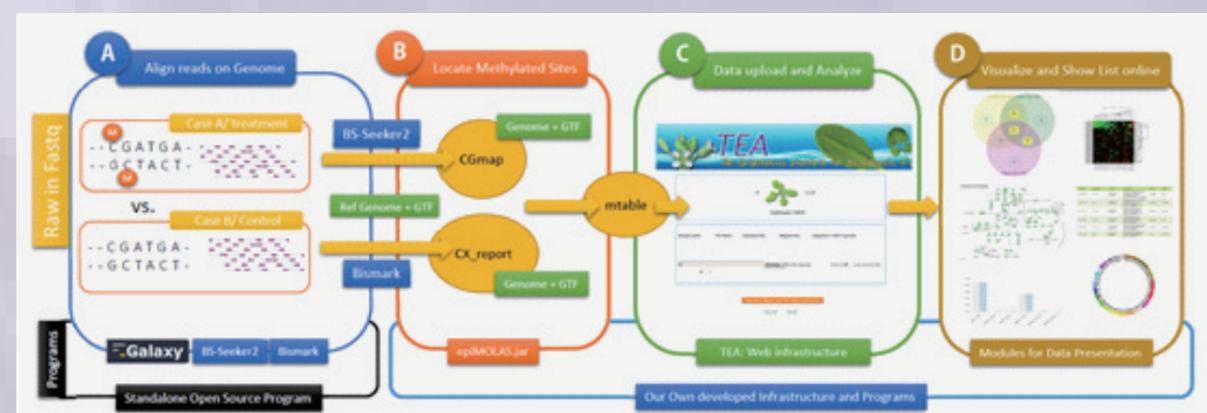
## 甲基化快速分析平台解析生物醫學 首創原始定序數據圖像化呈現完整分析平台

林仲彥  
副研究員

在二十世紀末基因體定序計畫開展之後，生命醫學研究發生了革命性的改變，其中最顯著的影響之一，就是我們開始能以全基因體的視角來研究生命的奧妙變化，而生物學的研究，也跨入了大數據時代。以人類為例，人類的基因體包含 23 對染色體與粒線體 DNA，以組成單元來計算的話，約為 30 億個以特定的方式排列的 DNA 鹼基對，也使得當代的研究者，必須隨著基因體大數據解析的浪潮，將研究的眼光由少數的基因標的，推升到可以鳥瞰全基因體的高度。在研究探索各種問題的歷程中，科學家們發現許多重要的問題並不能以序列差異（例如對偶基因）或發生異變（例如基因發生單點突變、染色體易位等）來解釋。舉一個最容易理解的例子：胚胎發育來說，我們的生命，始於單一細胞（受精卵），經由細胞分裂過程產生新的子細胞；子細胞雖然源於母細胞，卻不

見得會保持完全相同的模樣與功能。最後到出生的階段，人體已有數百種帶有相同的遺傳資訊而型態功能各異的細胞。針對同一個”劇本”（基因體），在不改變內容（鹼基序列）的前提下，如何微調控制基因體活動，以產生多種解讀方式的研究，就是表觀遺傳學 (epigenetics) 與表觀基因體學 (epigenomics) 的重點。而從原始大量數據處理到後續的基因調控解析，都需要資訊技術與生醫專業領域知識的投入與整合，才能從複雜的生命現象中逐步地抽絲剝繭，更深入我們所關注的研究主題。

本研究室長期投注在建構模式與非模式物種的多維體學 (multiomics) 分析平台，因此在基因體、轉錄體與蛋白質網路學上，有許多整合經驗。這個全基因體甲基化概況 (whole genome methylome) 分析系統的研究，是與中研院植微所陳柏仰老師實驗室的合作計畫，並獲得

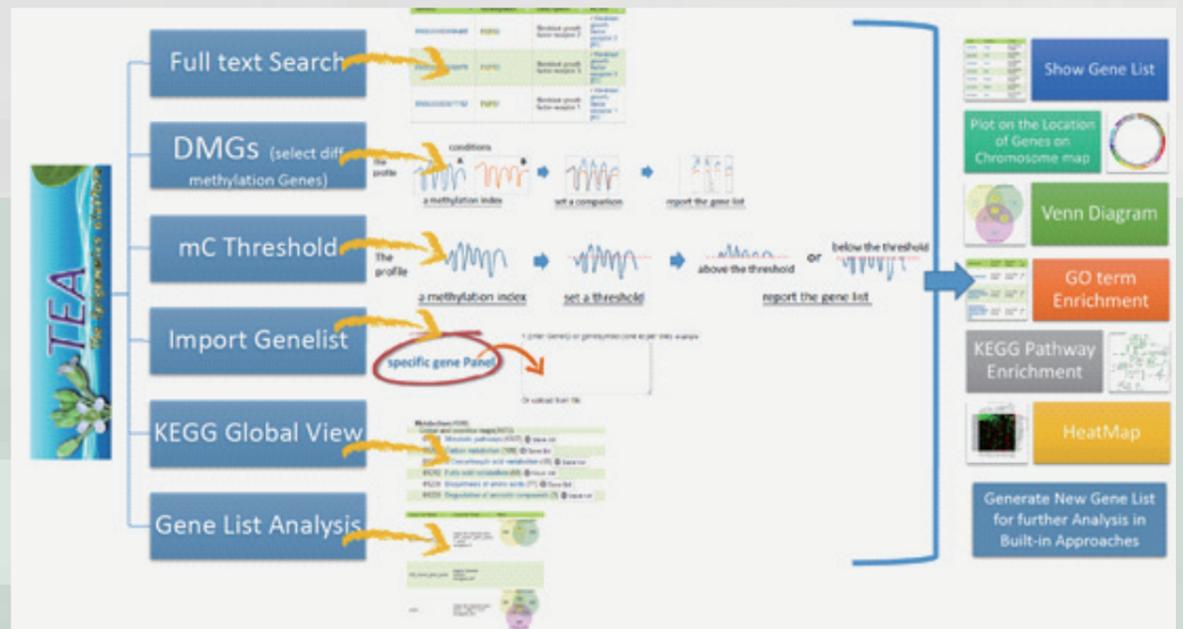


TEA 的運作流程。(A) 首先，使用特定的序列比對軟體將原始的 BS-Seq 讀序對映到參考基因體序列，(B) 把前述步驟所得的對映結果 (CGmap 或 CX\_report) 以 epimolas.jar 解析為 mtable，(C) 上傳至 TEA Server，(D) TEA Server 將上傳的資料自動佈建為分析網站。(圖一)。

科技部與微軟公司全球雲端研究計劃贊助。透過此一整合分析平台，結合陳柏仰老師所發展出來的甲基化偵測演算法 (BS seeker2, BMC Genomics 2013, <https://github.com/BSSeeker/BSseeker2>)，我們建構了新的分析流程，簡化從原始數據處理到甲基化差異位點找尋的複雜流程，並結合後續相關基因註解資訊，協助非專精於生物資訊工作的研究者，能以如同網頁瀏覽搜尋的操作感，以線上即時操作的方式結合其專業知識，來檢視全基因體甲基化概況的實驗結果，並進行後續的功能性分析，更深入地瞭解實驗過程中所影響到的甲基化變化，是如何來調控複雜的生理機制與基因互動。

## 研究重點

5-methylcytosine (5mC) 是染色體 DNA 中含量最多的非典型鹼基，這是在胞嘧啶 (Cytosine，縮寫為 C) 鹼基環上的第五個碳發生甲基化修飾，5mC 具有可傳遞、可被添加或消除的特質，因此成為表觀遺傳學與表觀基因體學研究的對象。在研究上可應用進行亞硫酸鹽反應，將未甲基化的 C 轉變為 U，利用定序或甲基敏感性限制酶切位變化，來檢測基因體是否帶有甲基化修飾。近年來由於高通量定序 (high-throughput sequencing, HTS) 技術的快速進展，研究者可以抽取基因體的 DNA，以亞硫酸鹽反應後進行全基因體隨機讀序，藉由分析未轉變為 T 的 C 以回推估算個別胞嘧啶發生甲基化的頻度，這使得我們在的 5mC 分布與變動研究上，可以達到全基因體、單一鹼基的解析度。5mC 與其他型式的鹼基代謝 (如 5-hydroxymethylcytosine, 5-hmC) 都可能具有特殊控制意義，是目前相當炙手可熱



TEA 的分析模組。利用全文檢索 (full text search)，或是 KEGG 代謝圖，使用者可以查閱個別基因的甲基化狀況，也可以用差異性分析或是閾值選取特定甲基化樣態的基因群，進行功能叢集分析，或繪製色溫圖。(圖二)。

的研究主題，也是相關多維體研究的重點之一，然而，全基因體甲基化實驗資料分析相當複雜，例如說，偵檢出的 5mC 的位置是坐落在基因的內部還是調控區？一個檢出的 5mC 的左右是否鄰近 G (形成特殊的對稱構造)？基因與調控區的甲基化程度要如何估算？要如何整合基因體甲基化的分析及轉錄體、基因體與代謝體等其他多維體的結果來解釋複雜的生物調控機制？對於無法進行大數據處理、生物資訊分析經驗有限或是無全基因體分析經驗的研究者來說，非常困難，全然是一個不可能的任務。

## 結果

為了將全基因體甲基化分析工作化繁為簡，本團隊開發了第一個從原始定序數據開始，到結果圖像化呈現的完整分析平台，讓研究人員能更專注在分析結果背後所代表的生物意義。

首先，我們以常見的模式植物，阿拉伯芥 (*Arabidopsis thaliana*) 為主體，建構全基因體甲基化分析平台，TEA (The

Epigenomics Platform for Arabidopsis, <http://tea.iis.sinica.edu.tw/>, Su et al., 2016, BMC Genomics), 其整體系統的規劃如圖一所示。分析流程上，首先將基因體亞硫酸鹽定序所得的隨機短序列，以特定的序列比對工具如 BS Seeker2 ([http://pellegrini.mcdb.ucla.edu/BS\\_Seeker2/](http://pellegrini.mcdb.ucla.edu/BS_Seeker2/)) 或是 Bismark (<https://github.com/FelixKrueger/Bismark>) 對映到基因體上，總結為每個 C 發生甲基化與非甲基化的計數數值，之後依據個別 C 的鄰近序列類型分為三類：CG, CHG, CHH，依位置 (基因本體與調控區) 結算其 C 甲基化程度，正規化為 0 (所有觀測 C 都未甲基化) 至 1 (所有觀測 C 都甲基化) 之間的數值，每個基因可產生六個甲基化程度估算值 (gene-CG, gene-CHG, gene-CHH, promoter-CG, promoter-CHG, promoter-CHH)。我們提 (文轉至第 11 頁)

## 程式語言與形式方法實驗室 邏輯與數學為工具確保軟體正確性

穆信成  
副研究員

程式語言與形式方法實驗室共有四位成員，其共同之處是以邏輯與數學為工具，運用形式化的方式確保軟體之正確性。近年來研究方向之一是與「單子 (monad)」有關。

純函數語言中的「函數」是真正的數學函數。因此擁有數學的好處：程式可當作數學來理解，可用包括歸納法在內的種種數學工具去推論關於程式的性質。

除了將輸入與輸出值對應外，電腦做的其他事稱作「副作用」。這包括更新變數的值（改變「狀態」）、拋出例外 (exception)、輸出入、非確定性 (non-determinism)、並行 (concurrency)... 等等。一般的印象是函數語言中不允許副作用的存在。事實上，許多實用的函數語言甚至支援許多種副作用。只是在支援副作用的同時不能失去原有的好處：副作用也必須以「能以數學方法管理」的方式出現在程式中。其中一種管理、描述副作用的方式便是透過「單子」。

一個單子包括一個型別建構元  $M$ ，和兩個運算子 `return` 和  $(>>=)$ 。含型別  $Ma$  的值代表「一個『計算』，若執行後將傳回型別為  $a$  的結果。」在執行該計算的過程中可能會發生副作用。算式 “`return x`” 表示「直接傳回  $x$ 」（沒有其他副作用）；“`m >>= f`” 則表示「將執行單子  $m$  的結果傳給函數  $f$ 」（後者將傳回另一個單子）。它們須滿足三條「單子律」：

例如，左同一律表示「把『直接傳回  $x$ 』的結果丟給  $f$ 」和「把  $x$  丟給  $f$ 」是一樣的。結合律則讓我們改變  $(>>=)$  結合的優先順序。除此之外，一個單子可能支援一些副作用。每個副作用都需要一些額外的運算子。如果我們想要非確定性，通常會假設兩個運算子  $mzero :: Ma$  和  $mplus :: Ma \rightarrow Ma \rightarrow Ma$ ，前者表示「失敗」，“`x 'mplus' y`” 則表示「本計算的結果可能是  $x$  或  $y$ 」。在合理的情境中， $mplus$  應該有結合律， $mzero$  應該是其單位元。如果想要有狀態，通常會假設有兩個運算子 `get` 和 `put`，可分別讀取和覆寫一個只存在該單子中的變數的值。它們也有幾條應滿足的規則。這些規則可用來論證、推導程式的性質：即使有副作用，程式的性質也是我們所能掌握的。

將單子的觀念引入計算科學中可追溯至 90 年代初。當時，單子是新話題，最令人好奇的自然是「怎麼實作」？支援非確定性的 `return, (>>=)` 有幾種寫法，支援狀態的 `return, (>>=)` 也有幾種寫法，進行研究時最初在意的可能是他們的效率。

很快地發現到，多種副作用的單子很難彼此組合。例如，同時支援非確定性和狀態的 `return, (>>=)` 很難從現有的推論出來。難道每加一個副作用，就得重寫一次嗎？95 年有人提出「單子轉換器」，可想成從單子到單子的函數，可以直



接結合。此時較關切的是單子的模組化等較工程、「實用」的問題。

後來有許多意外的發現。多數人以為 Maybe 型別可當作一種較有效率的非確定性單子，後來發現它並不滿足一條重要分配律。最直覺的非確定性單子轉換器寫法作用在狀態單子上，轉出的東西連單子律都不滿足。最近才聽說，非確定性和機率（另一種單子）兩者單獨存在時都滿足該有的規則，同時出現在一

個語言中時卻不再滿足所有單子律了。

因此近年來談單子的方式又有所轉變，其方向在給定一些性質，如何做出滿足這些性質的單子？另一方面有人認為單子轉換器可能不是好作法，而提出其他可能。但我們則研究如何證明單子程式的性質，並嘗試釐清確保多種副作用結合後仍保持其應有性質的充分條件。

## 甲基化快速分析平台解析生物醫學

（文轉自第 9 頁）

供了一支程式 `epimolas.jar` 來做這些計算，規範了一個輸出格式並設計了簡潔的上傳與資料庫佈建流程，透過網路介面，研究者可將甲基化定序資料解析報表上傳至我們的主機，連結系統內建的阿拉伯芥基因體註解資訊，包含個別基因的註解資料庫和數據分析平台，即可建構一個阿拉伯芥基因體甲基化分析網站。在阿拉伯芥基因體註解部分，我們維持在目前最廣泛使用的 TAIR10。我們也設計了上傳資料所建構的資料分析網站的分享機制，以使用者的設定來決定上傳資料為公開分享，或以密碼分享於研究團隊中，或僅供個人使用（圖一）。

透過 TEA，研究者可以進行數值分析，由甲基化程度的差異，或是設定一個閾值，挑選出符合條件的基因集合，進一步以功能傾向來分析基因集合的生物意義，或是將這些基因的資料以叢集色溫圖表現，或標定出基因的位置。此外，

也能夠以關鍵字與基因集的搜尋方式，列出特定基因或是一小群基因的甲基化程度估算值（圖二）。

### 結語

生物資訊研究不斷演進，隨著新世代的實驗方法的演進，而造成生物序列巨量資料的快速累積，隨之而來挑戰包括數據傳輸、計算解析及儲存應用等，已跳脫原有的生物醫學領域，需要資訊技術的深入協助。目前，我們正面臨資訊潮流的快速衝擊，而巨量資料所帶來豐富特性，將能結合深度學習等新一代人工智慧技術，提供我們挖掘這些數位寶藏的種種可能性。TEA 是我們與植微所陳柏仰老師合作，所建立阿拉伯芥基因體甲基化分析平台，目前在人類與小鼠的基因體甲基化分析平台也已建構完成，正進行論文的撰寫與雛形測試。同時，我們也將甲基化前端原始數據處理

的複雜流程，建構簡化為 GALAXY 平台上的分析工具，未來將透過 DOCKER 輕量級虛擬技術的協助，提供跨平台的應用。在過去的幾年內，我們建構整合上下游分析流程發展生物資訊平台，其意義乃在於透過清晰易懂的使用者介面，結合使用者的專業知識，來協助他們進行後續生物意義分析與圖像化呈現，而長程目標是在於資料與分析方法整合與可視化，包括轉錄體、基因體甲基化等蛋白質交互作用網路的整合。我們希望能結合各種實驗方法所產生的不同維度觀測，打破生命科學的巨量資料分析瓶頸，使多維體分析更為直觀。

### 參考資料

1. Sheng-Yao Su, Shu-Hwa Chen, I-Hsuan Lu, Yih-Shien Chiang, Yu-Bin Wang, Pao-Yang Chen, Chung-Yen Lin\*, "TEA: The Epigenome platform for Arabidopsis methylome study," BMC Genomics, volume 17(S13), pages 1027, December 2016, Tea website: <http://tea.iis.sinica.edu.tw>



## 法官們的困境 多數決對嗎？

蘇哲平 / 廖純中

**群**體的內部有意見上的衝突時，有時我們會訴諸投票，以多數意見做為集體意見。本文想介紹一個關於多數決投票的謎題——「法官們的困境」(Doctrinal Paradox)。藉由反省這個謎題，我們可以觀察到多數決的一個瑕疵。

### 法庭上的三位法官意見分歧…

在一個法庭上，有三位法官正在共同審判一個違約案件。這個案子的關鍵點有二：

1. 依合約規定，被告是否被禁止做某特定之 A 行為？
2. 被告是否真的有做 A 行為？

如果上述兩個問題的答案皆為肯定，那麼被告違約成立。但是如果有一問題的答案為否定，被告的行為不構成違約。

很不巧地，三位法官的意見分歧。三位法官的各自的意見如圖一所示。於是三位法官想透過多數決，來解消意見上的衝突，以多數人的意見做為集體意見。

### 一個形成多數意見的流程

對於被告有無違約，有一個形成多數意見的流程如下：

- 步驟一：三位法官先各自形成自己的結論(如圖二所示)；
- 步驟二：然後，再針對法官們各自的結論進行多數決(如圖三所示)。

三位法官中有兩位的結論為，違約不

成立。所以，針對所有法官各自的結論進行投票，結果會是：被告違約不成立。

### 另一個形成多數意見的流程

有趣的是，關於被告是否違約，有另外一個形成多數意見的流程：

- 步驟一：三位法官先針對案件的兩個關鍵點，進行投票，藉此形成集體推理的出發點(前提)(如圖四所示)。
- 步驟二：再基於這些集體推理的出發點(前提)做推論，判斷結論為何(如圖五所示)。

對於案件的一個關鍵點——依合約，被告是否被禁止做 A 行為——多數意見為肯定。對於另一個關鍵點——被告是否真的有做 A 行為——多數意見也為肯定。所以，在針對這兩個關鍵點投票後，三位法官以下面兩個陳述做為集體的推理出發點：

- ◆ 依合約，被告應被禁止做 A 行為；
- ◆ 被告真的有做 A 行為。

基於步驟一得到的集體推理出發點(前提)，可以推出：被告違約成立。三位法官以此做為集體的結論。

### 法官們的困境

直覺上，上述兩個流程似乎是一樣合理的。但是，兩種流程各自得到的不同的「多數意見」：第一個認定，違約不成立；第二個流程則認定，違約成立！那到底法官們應該判，被告違約成立或不成立？這個問題似乎並不好回答。

	依合約不得 做A行為	有做A行為	違約
法官甲	YES	YES	
法官乙	YES	NO	
法官丙	NO	YES	
多數意見			?

圖一

以上即為「法官們的困境」謎題。圖六統整了整個謎題。

### 這個謎題的啟示

「法官們的困境」告訴我們什麼？要回答這個問題，我們需要先對這個謎題進行簡化。

首先，讓我們：

- ◆ 忽略圖六中所有橫向的箭頭；
- ◆ 暫時「清空」圖六中，多數意見的欄位。

這時我們會得到圖七。在這裡，我們所保留的是，每位法官對三個問題，各自的判斷。接著，讓我們觀察，對這三個問題進行多數決的結果。（如圖八。）值得注意的是，圖八是圖六的一個簡化。在圖八，我們省略了所有橫向的箭頭。仔細看簡化後的圖表八，我們可以發現兩件事：

- ◆ 每位法官個人的立場，都沒有自我矛盾。
- ◆ 但是，針對這三個問題，所形成的多數意見卻是自相矛盾的。多數意見認為：
  - 合約有禁止 A 行為；
  - 被告有做 A 行為；
  - 違約不成立。

前面兩個陳述加起來與第三個陳述矛盾，因為從前面兩個陳述，我們可以推出違約成立。

所以，「法官們的困境」(Doctrinal Paradox) 這個謎題告訴我們：當我們針對複數個（有相關聯的）議題進行投票時，其多數決的結果可能會是自相矛盾的！

	依合約不得 做A行為	有做A行為	違約
法官甲	YES	YES	YES
法官乙	YES	NO	NO
法官丙	NO	YES	NO
多數意見			

圖二

### 是否可以逃脫困境 來自於人工智慧的協助

以上的說明中，法官們在對某個議題進行投票時，是採用多數決的辦法，雖然這是民主社會常用的方法，卻不是唯一的方法。事實上，多數決只是眾多我們稱為整合方法之中的一種而已，在此處，因為其係用於整合法官們對於議題的判斷，因此也稱為判斷整合。在社會選擇理論中，諾貝爾經濟學獎得主 K. Arrow 早已證明，在某些合理條件要求下的偏好聚合方法是不存在的。同樣的，要找到能滿足一些基本合理要求的判斷整合方法，也已經被證明是不可能的。然而，法官們的困境是否無法逃脫呢？為了探討這個問題，學者們對於人工智慧中用於多代理人信念融合的技術產生了興趣。由於每個代理人從環境中收集的資訊有所差異，因此可能形成不同的信念，就如同每個法官對同一個議題可能會有不同的判斷一樣，將所有代理人的信念直接融合在一起，可能會有邏輯上的矛盾產生，因此人工智慧的研究者便想出一種從語意上著手的方法。其主要概念是，每個個體的信念可以對應到其所滿足的語意模型，也就是，

	依合約不得 做A行為	有做A行為	違約
法官甲	YES	YES	YES
法官乙	YES	NO	NO
法官丙	NO	YES	NO
多數意見			NO

圖三

	依合約不得 做A行為	有做A行為	違約
法官甲	YES	YES	
法官乙	YES	NO	
法官丙	NO	YES	
多數意見	YES	YES	

圖四

	依合約不得 做A行為	有做A行為	違約
法官甲	YES	YES	
法官乙	YES	NO	
法官丙	NO	YES	
多數意見	YES	YES	YES

圖五

	依合約不得 做A行為	有做A行為	違約
法官甲	YES	YES	YES
法官乙	YES	NO	NO
法官丙	NO	YES	NO
多數意見	YES	YES	YES & NO

圖六

(文轉至第 16 頁)

## 空氣盒子空污偵測新銳

副研究員 陳伶志



中研院是台灣學術研究的最高殿堂，在我立志走上學術研究這條道路時，自然而然以中研院資訊所為第一目標。於是，於博班最後一年，抱著孤且一試的心態送出申請書，驟然收到當時網路組召集人郭譽申老師的回信，並且邀請我回台灣演講介紹當時研究的主題。彷彿記得那場演講吸引了為數不少的教授，不久之後，所方便通知我，資訊所有意願延聘加入研究團隊，喜出望外之際於是展開準備畢業的程序，在我畢業典禮結束之後，很順利的便成為資訊所的一員了。

### 緣起

我的研究領域其實大方向一直沒有改變，但在比較具體的題目選擇上已經改

了好幾次。大方向沒變，是因為我從學生時代就一直對實用性高的問題感興趣，很喜歡先動手熟悉既有系統的原理，然後找出不滿意的地方，接著就動手去修改，或者加入自己設計的功能。我記得念博班時的第一個題目，就是在學完無線藍芽的技術標準後，發現藍芽技術雖然有許多的通訊模式，但並無法根據網路狀況，動態切換到效率最高的工作模式，因此我就設計了一個自我動態調適的演算法，並且放進藍芽模組中，後來這個研究不但發了我人生第一篇論文，這個概念在後來藍芽發展中，也變成標準中的一部分了。

我的博班題目主要也是針對網路頻寬量測，針對當時非常耗費資源的量測方法，從理論分析開始，設計又快速又準確的新方法，後來又針對一連串不同的網路連線技術，設計各種適用的量測方式，這些開發出來的方法，後來都有做開源，並且被廣泛使用，也由於這次開源的經驗，讓我體會到讓更多人使用到自己研究的成果所帶來的成就感。

### 個人簡述

加入資訊所團隊後，我的研究題目開始轉向無線感測系統，第一個建立的大型系統是在玉山國家公園建立的登山客資訊系統。在那個系統中，我們用「接力式」的短距離無線網路傳輸，突破山

區無法建立高覆蓋率無線網路的困境，並且針對登山客的需求，量身訂做體積小、功率低且具備網路連線功能的「登山客黑盒子」。這個計畫持續了近四年的時間，從發想到實作，爬了不少小山做系統測試，到最後爬了好多次玉山進行系統佈建，這是我第一次把研究場域從實驗室移到戶外場域，中間當然也繳了不少學費，從失敗中學到不少寶貴的經驗，但是在解決問題的同時，看到自己開發的系統，被應用在真正的現實問題中，並且可以解決迫切需要的問題，那種成就感是難以用言語描述的。

在玉山登山系統之後，我的下一個系統作品是公車舒適度量測系統，利用智慧手機內建的三軸加速度計，測量乘客在搭乘公車時手機隨著車子開動所承受的振動狀況，並且套用一些理論進行舒適度的評分。接著我們又利用手機上 GPS 的紀錄，結合公車即時資訊系統的開放資料，進一步找出公車的各項屬性，並且在收集大量資料後進行整個城市規模的公車舒適度評鑑系統。這個系統剛好搭上了智慧手機和開放資料的風潮，因此當時算是很創新的一套系統，也曾得到一項大獎，但也因為真的走在太前端，過度仰賴使用者自發性的參與，卻沒有考慮到使用者經驗層面的議題，因此後來很可惜在推動上一直無法有重大突破，無法真的做到原本預設的佈建規模，但

從這之間所學到的教訓，卻是後來進行系統研究的寶貴經驗。

## 研究目標及內容

我目前的研究主要聚焦在以群眾參與為基礎的大型PM2.5感測系統，簡稱「空氣盒子專案」。這個研究其實已經默默進行超過五年了，但是直到最近這兩年才忽然有突破性的發展，在前三年幾乎都是在繳學費學經驗的過程，甚至已經到了快要放棄的狀態。之所以在這兩年有快速的進展，一方面歸因於感測器技術的突破，讓量測的結果越來越有代表性，一方面則要歸因於這幾年國內創客運動的興起，以及網路上各種公民科學社群的活躍，當然，這幾年台灣空氣品質日漸惡化，讓民眾越來越重視空污問題，也越來越願意參與相關的公民科學活動，也是很大的原因。

目前空氣盒子的計畫已經越滾越大，從一開始兩位數的佈建數目，在半年內達到三位數，一年內達到四位數，預定在兩三年後將會突破五位數，而在參與的國家方面，在不知不覺間也已經突破30個國家了，希望今年底可以再有倍數以上的成長。雖然表面上這是一個硬體佈建的計畫，但是背後其實隱藏許多紮實的研究當根基，隨著佈建規模不斷的擴大，更累積越來越多新的研究議題，這些議題有些非常工程，有些非常學理，但有一個共通的特性，就是這些議題都非常的實際，同時也是非得要先有一個大型的系統佈建，否則絕對無法空想出來的。目前的研究工作也隨著這些議題，

變得越來越忙碌，但是每天都充滿新的挑戰，每天都覺得充滿了戰力，特別是看到問題一個個被解決的時候，其實是讓人非常興奮的。

## 對未來資訊所及資訊研究的願景

資訊所的教授每位來自世界頂尖，每個人在其研究領域都非常傑出。而目前資訊所的研究方向中，我所從事的的研究偏向於實用性導向，有時不免感到孤單，但更多時候是感到資訊所資源非常豐沛，想找任何一個領域的人諮詢，都可以很容易就近找到這個領域的專家，因此合作上其實反而很容易激盪出很棒的想法。不過這幾年整個資訊領域在物聯網、網路通訊技術、資料科學還有人工智慧等議題的進展真的太快了，這部份確實是需要大家一起努力的，才能讓資訊所繼續走在全世界資訊研究的尖端。

## 對未來想從事該領域研究的後進有何建議

這幾年來，常常感覺到目前國內資訊科系訓練出來的學生，在實作能力上有越來越沒有動手能力的趨勢，一方面可能是現在的電腦系統真的太方便了，所以也把大家都寵壞了，一方面也有可能是因為現在新技術新領域發展的速度太快了，所以大家都習慣等待最後的結果，以及別人分享出來的資源與工具。最後導致的結果就是，現在多數的學生普遍少了自我摸索的能力，甚至畏懼探索未知的領域。更糟的是，很多學生開始看

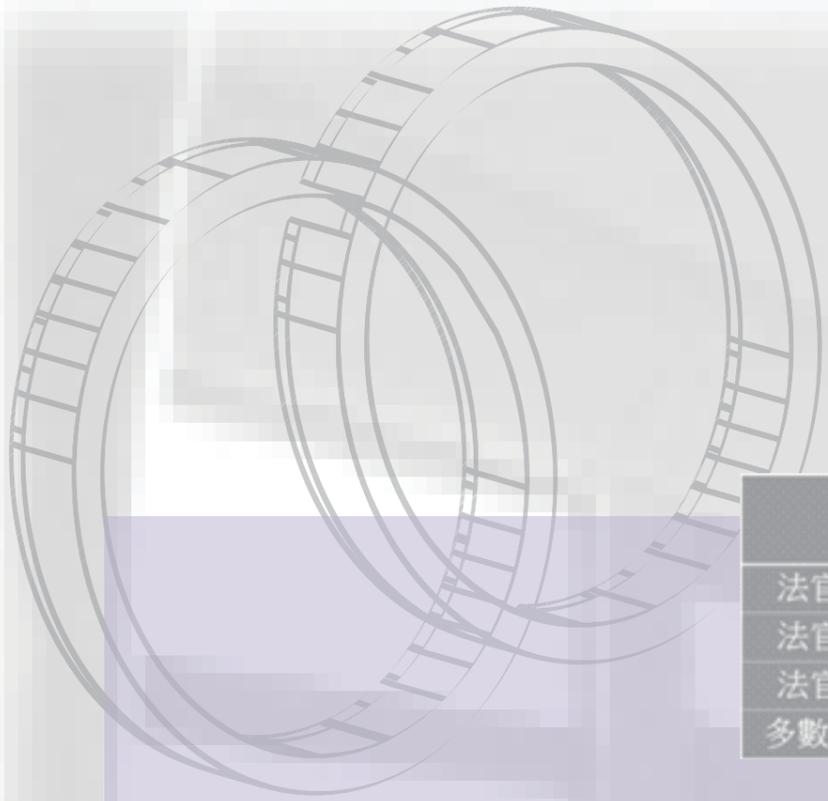
輕動手做的系統能力，只一味追求可以很快看到成果的題目。但是，「天下難事，必作於易；天下大事，必作於細」，這些動手實作還有自我探索未知領域的過程，雖然不一定每一次都會有好的結果，但卻是累積自己能力的最好機會，對於未來從事資訊研究，更能增加觸類旁通以及創新發展的能力與機會。

## 生活及休閒活動

其實生活除了研究之外，大概就是以家庭為主，特別是陪伴兩個兒子，我很享受和他們相處的時間，雖然有時候真的很累人，不過只要看到他們開心的樣子，就能把一切的疲累完全忘掉，有時真的忙的時候，也會帶他們到辦公室加班，一邊讓他們玩積木，一邊自己做自己的工作。除此之外，其他的時間就是喜歡學一些新的技術，甚至自己動手玩一些新的工具，例如目前研究的空氣盒子，其實最早也是自己私下亂玩的產物。

## 對個人自我未來的期許

除了把學術研究工作做好之外，對自己最大的期許，就是做出大家覺得有用、好用，並且常常會用、不用不可的系統，我一直相信科技始終來自於人性，也希望自己可以一直保有從事資訊研究的初衷，在從事研究的路上，可以不斷做出自己滿意的成果。



	依合約不得 做A行為	有做A行為	違約
法官甲	YES	YES	YES
法官乙	YES	NO	NO
法官丙	NO	YES	NO
多數意見			

圖七

## 法官們的困境 多數決對嗎？

(文轉自第 13 頁)

一個個體的信念代表的是一組可能的世界圖像，因此要找出眾人信念的融合結果，就是去找到距離眾人對世界圖像的描述最接近的可能世界。由於在有限議題下，這樣的可能世界一定存在，因此必能產生一種一致的整合結果。然而，由於距離最近的可能世界有時候不只一個，因此所產生的整合結果可能是對某些議題懸而不決，換句話說，法官們雖然逃脫了彼此衝突的困境，卻也陷入了無法判決的困境。因此，解決這個問題的方法，雖然因人工智慧的研究而露出了一線曙光，然而最終是否能找到一個

令人滿足的方案，還是有待未來進一步的研究了。

### 延伸閱讀

這個謎題可以推廣為多一般的現象？是否有投票方法，其結果一定不自相矛盾嗎？這是社會選擇理論 (Social Choice Theory) 研究的課題之一。有興趣進一步了解的讀者，可上史丹佛大學線上哲學百科 (Stanford Encyclopedia of Philosophy, 網址如下：<https://plato.stanford.edu/>)，搜尋 Belief Merging and Judgment Aggregation 或 Social Choice Theory 等條目。

	依合約不得 做A行為	有做A行為	違約
法官甲	YES	YES	YES
法官乙	YES	NO	NO
法官丙	NO	YES	NO
多數意見	YES	YES	NO

圖八



中央研究院資訊科學研究所  
115 台北市南港區研究院路二段 128 號  
發行人：許聞廉  
編輯：田力如，曾慧琦  
tel: +886-2-2788-3799  
[www.iis.sinica.edu.tw](http://www.iis.sinica.edu.tw)

