



國立交通大學
National Chiao Tung University

出國報告（出國類別：學研訪問）

赴美訪問史丹福大學王永雄院士及其研究團隊 出國報告書

服務機關：理學院統計學研究所、
丘成桐中心

姓名職稱：盧鴻興 院長、
阮鴻輝 博士後研究

派赴國家：美國 舊金山 史丹福大學

出國期間：102/02/16~02/21

報告日期：102/03/13

摘要

此行與理學院盧鴻興院長一同前往美國史丹福大學，目的是訪問王永雄院士及其研究團隊，自 2012 年夏季開始，交通大學團隊與王院士團隊已開始一連串的研究合作，針對王院士團隊開發的「無母數機率密度估測演算法」進行軟硬體加速，雙方合作甚歡，未來將持續進行研究合作，研究項目是針對巨量資料(Big Data)的統計分析/資料探勘之演算法暨軟硬體協同開發，以及此研究計畫的實質應用，盧鴻興院長在此研究合作計畫上是負責統計理論與生醫影像相關之應用，本人隸屬於盧鴻興院長的研究團隊，在丘成桐中心屬生醫影像小組，此次與盧鴻興院長訪問王永雄院士及其研究團隊，是就雙方研究合作內容進行交流與討論未來合作方向。

目次

一、目的.....	4
二、過程.....	5
三、心得及建議.....	8

一、目的

交通大學方面與王永雄院士研究團隊的研究合作，自 2012 年夏季開始，王院士團隊開發的--無母數機率密度估測演算法，名為貝氏循序切割(Bayesian Sequential Partition)演算法，此演算法在機率估測的效能上甚為卓越，特別是在高維度的資料下比起其他演算法能得到更準確的估測結果，然而此演算法若以純軟體版本實現，針對一般的資料量即需要顯著的時間方能執行完畢，因此需要軟體版本的加速，讓此演算法能得到更大的發揮空間，甚至用在巨量資料(Big Data)的分析上，交通大學在晶片設計與硬體實現上有很豐富的經驗，和王院士團隊在這方面的需求相符，自 2012 年開始，啟始了合作，使用「現場可程式邏輯門陣列(Field Programmable Gate Array, FPGA)」與「圖形處理器(Graphics Processing Unit, GPU)」的方案針對貝氏循序切割演算法進行加速，並預計今年五月會下線晶片。

基於先前的合作結果與經驗，王院士將目標放在更深遠的巨量資量的統計分析/資料探勘之上，將與交通大學團隊繼續合作，盧鴻興院長在此研究合作計畫上是負責統計理論與生醫影像相關之應用；電子所李鎮宜教授是負責演算法之硬體設計方案與矽晶片的開發；電子所王聖智教授是負責與電腦視覺、影像處理相關的演算法開發與應用；電子所張錫嘉教授負責記憶體相關技術及系統架構設計；電子所賴伯承教授負責 GPU 加速以及系統架構設計。此行與理學院盧鴻興院長出國前往美國舊金山，目的是訪問王永雄院士及其研究團隊，進行近期合作內容的交流，以及未來合作方向的討論。

二、過程

二月十七日至十八日，與王院士研究團隊討論貝氏循序切割演算法之軟硬體加速方案，包含現場可程式邏輯門陣列(後文以 **FPGA** 簡稱)加速版本與圖形處理器(後文以 **GPU** 簡稱)加速版本，以及五月的晶片下線之構想，此外，討論有關貝氏循序切割演算法之相關資料結構設計，以及機率密度估測結束之後，後端演算法應用時，存取新資料點對應之機率密度所需之貝氏循序切割樹狀結構之資料結構設計。

二月十九日星期二，本日雙方團隊進行研究交流，以正式簡報報告形式呈現，每段講題約數十分鐘至一小時。在早上的講題是交通大學電子所張錫嘉教授進行記憶體相關技術的簡報，敘述現今各種記憶體技術的優劣點與未來發展趨勢，以及各種記憶體技術在本次洽談之研究合作主題：『巨量資料的統計分析與資料探勘之演算法與軟硬體架構協同設計』上的適用情形之分析和評估。此外，王院士的終極目標是想達到聰明記憶體(**Smart Memory**)的記憶體內處理(**In-memory processing**)，與一般認知的記憶體內處理的不同點在於，一般認知的記憶體內處理是指資料運算時還是從儲存元件搬移到速度較快的記憶體上來準備進行運算；然而王院士的目標是想將基本的一些運算功能，能分散式地分布在儲存元件上來做，也就是負責儲存資料的裝置將能負擔一部分簡單的運算，例如計數(**counting**)、模式(**pattern**)比對，邏輯/集合式的計數等等，雙方就此想法的可能性，以及採用的記憶體技術，控制元件的評估，進行討論。

中午的部分開始，是由王院士團隊的三名成員進行簡報，第一位是吳東祐

先生，之前他是交通大學電子所的學生，畢業後曾在盧鴻興院長研究團隊擔任研究助理，平時王院士團隊與交通大學團隊在越洋視訊會議上的交流與討論也是他在負責，因此他報告的內容是與交通大學團隊討論有關巨量資料的彙總，包含王院士團隊想要達到的 **Smart Memory** 之相關需求，以及與交通大學團隊初步評估系統架構與記憶體架構分析的結果。第二位是貝氏循序切割演算法的作者之一 **Ms. Lu**，她的講題是有關貝氏循序切割演算法在資料壓縮上的應用，包含無損壓縮以及有損壓縮兩種模式。第三位 **Yang** 報告的講題是在貝氏循序切割演算法在生物資訊上的應用，當貝氏循序切割演算法估測完機率密度後，後端作分群的演算法之結果。王院士最後也就基因資料(**Genome data**)的內容進行一些介紹，包含他們想要做的在 **Genome data** 這樣的巨量資料上的一些分析方法。

下午的部分，由交通大學團隊報告在交通大學這邊就雙方合作上的構想書簡報，講題是『**NCTU Intelligent Big Data (iBD) Research Center**』，簡報者為李鎮宜教授與盧鴻興院長，報告的內容包含交通大學五位教授研究團隊的研究提案，盧院長研究團隊在交通大學團隊的分工上，負責項目包含統計分析、方法論與演算法開發，以及在生醫影像如超聲波影像、**MRI** 影像、神經影像上的應用，盧院長報告的部分是與丘成桐中心生醫影像處理小組相關的部分。

交通大學團隊報告結束之後，王院士與盧鴻興院長、李鎮宜教授繼續討論未來雙方合作模式以及細節；王院士的博士後研究員則是帶著交大參與研究團隊的成員參觀王院士的實驗室與儀器設備。參觀結束之後，交大研究團隊成員繼續跟王院士研究團隊的 **Ms. Lu**、吳東祐等人討論貝氏循序切割演算法的軟硬體加

速設計與演算法應用等細節，例如應用在生醫影像上的分群、影像切割、模式辨認(Pattern Recognition)等等。

此外，由於交通大學下學期已開學，盧鴻興院長星期四有課，故盧鴻興院長在本日開會結束之後，搭機先行返臺。

在二月二十日至二十一日，與王院士團隊的吳東祐、Ms. Lu 進行討論，包含貝氏循序切割演算法之 FPGA 加速版本與 GPU 加速版本如何讓他們可以使用，以加速資料分析效率，還有交通大學團隊五月的晶片下線之細節規畫，包含貝氏循序切割演算法以及前端資料特徵擷取、後端演算法應用的規畫等等。討論結果暫定是，此次晶片下線會將整體貝氏循序切割演算法加上後端的分群/分類演算法實現在硬體上，下個階段則是考慮將 Ms. Lu 的基以貝氏循序切割之資料壓縮演算法也實現在硬體上。

三、心得及建議

不論是在美國或是全世界，巨量資料(Big Data)的分析是十分熱門的新科技。巨量資料在目前已經隨處可見，像是健保病例、社群網站資料、基因資料、生醫影像、財務資料等等，如何分析巨量資料以及如何建立一套完整個系統架構來實現，不僅在學術上有所貢獻，在商業上的價值或者是對全人類全體福祉改善上都有所貢獻。未來的研究題目，趨勢可能是資料越來越大量、題日本身越來越複雜，需要跨領域的團隊合作，才能完成。這次能和史丹佛大學王院士合作交流，是難能可貴的機會，未來交通大學可以把握自己的長處，多尋找跨領域交流的機會，在校內可以先構築一跨不同研究領域的團隊，並有初步的一些結果與貢獻，在尋找校外合作機會時會比較容易成功。若能找到與美國一流大學的合作機會，是一個首選。因此王永雄院士、國科會駐舊金山科技組組長、盧鴻興院長、李鎮宜教授、與參與團隊成員在二月十九日餐敘時熱烈討論，大家都願意積極地促成交通大學與史丹佛大學王院士團隊雙方的合作交流。