

文組生與程式的距離 ——從零開始寫程式的想法與工具

蘇洪寬 國立臺灣師範大學英語學系語言學組博士生

前言

這篇文章希望能給想開始接觸程式卻猶豫著的文組大學生以及其他讀者，一些想法以及可以善用的工具，相信在這個 AI 呼聲極高的時刻，在大學校園中，我們時常看到「AI 與某某領域結合」的相關演講，相信絕大數的文組生會如同當初的我，想著這應該是資訊專長的學生有興趣的內容，因而低頭加速離去，又或是覺得裡面感覺牽扯到極為複雜的程式，遠遠不是一個已經被必修、選修搞得焦頭爛額的大學生所能處理的。透過這篇文章，我希望帶來不一樣的看法。AI 其實比我們所想像得還要容易上手，只要我們願意去嘗試認識它，就會發現它有趣的一面。更重要的是，我們也能運用這樣的技術在相關的人文領域，進而協助我們的研究、或是驗證我們的觀察。

這篇文章也同時是我個人的學習歷程，從一個排斥程式的大學生，慢慢地能夠看懂一行行程式碼，最後更能因應自己的研究來撰寫合適的程式，很多同學曾問我：「從大學，更甚是研究所才碰程式是不是很辛苦？」但是在學期過程中，我們一定都會上過自己在之前都沒有經觸過的領域的課，當時的我們會不會覺得辛苦？抑或覺得有趣而樂於探索呢？如果是後者，那又是為何

麼？細細想來，無非是這個新興領域雖然陌生，但是我們知道其與我們的興趣有著千絲萬縷的關係，而透過學習這個新的知識或是理論架構，我們能更清楚地了解特定的現象。

一樣的道理也可以應用在程式學習中，如果僅單純學習程式語法，而缺乏跟自己領域的結合，很容易就因為缺乏即時的回饋而逐漸磨耗對於課程的耐心，這時我們應該運用課後的時間去進行額外的練習，去看看自己平常看的文章裡面的結果有哪些其實是可以透過程式畫出來，又或是平時看的研究有哪部分可以用比較量化統計的程式分析方法，當我們心中有「哇，原來平時我看到的資料可以這樣做！」或是「這樣做感覺也蠻有趣的！」的時刻時，我們在學習上就會更有動力跟目標，總而言之，找到與自己原本領域的關聯性，便是使我不停在程式這條道路上鑽研的動力。

一些想法：一個文組生怎麼開始接觸程式？

我接觸程式的起始點其實相對較晚，是到研究所才開始正式接觸到量化分析研究，還記得當時我們的選課清單上出現了 R 統計分析、程式語言用於語料分析導論、語料庫語言學等陌生的課

名。當時我剛進碩士班，一切都十分陌生，我在師大英語系語言組就讀，本來預期的課程是一些以前在書上常讀到的音韻學，句法學，構詞學等等，但是為什麼會出現程式課程？我當時十分疑惑，而且我那時完全沒有學過任何一點程式，因此這些課程名稱看起更令人退避三舍，不過當時在我的恩師陳純音老師的極力說服下，我才硬著頭皮選了一門程式語言用於語料分析的課程，還記得當時老師跟我說：「學習不應該偏食而是要多方嘗試，質化跟量化方法應該都要並重。」這就是我第一次跟程式語言結緣。

還記得第一次上程式語言用於語料分析導論時，看著一旁似乎本來就有底子的同學駕輕就熟地打開指定的程式，「哇！他的是黑色的介面，一定是很資深的那種」，望著我還是一片白色的軟體介面，心中的不安又加上了一層，這明明是完全不影響程式結果的東西，但是對於一個程式新手而言卻也變成了一種壓力。這是十分常見的現象，未知會帶來額外的挑戰以及抗拒。不過當授課老師慢慢從頭開始介紹，我的壓力漸漸被一個一個能順利完成作業的喜悅所消除，但是真的讓我產生動力主動去學習的契機是始於課堂的後期，老師拿出語言學裡頭蒐集到的實際聲音資料，讓我們試著用所學的圖表去呈現複雜的表格資料，當我看到原本錯總複雜的表格，在程式中視覺化之後所得到的有趣結果，與我們在語音學上所學的觀察有呼應之後，我開始感受到程式與所學之間的連結，也開始知道自己所學是真的有幫助的！因此如果你已經走出了第一步，鼓起勇

氣報名了一堂程式課程，一定要撐過前面的基礎語法學習！（通常是前六週），並同時尋找你目前的所學跟你專業之間的連結，這才是最重要的。

一些工具：閱讀以及讀書會

當我發現學科跟程式該如何連結之後，我便對程式語言的學習充滿動力，也開始主動尋找相關的資源來補充自己的知識，在這裡我想推薦幾本當初一路學來覺得很有幫助的書，希望也能同樣幫助到需要的同學們。

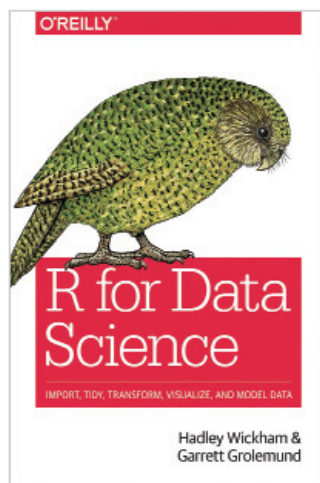
在此之前，我想先簡單談談我整個學習的發展架構：一開始因為我從課堂中學習到的，便是運用程式整理資料以及繪圖的方式，因此剛開始的閱讀範圍主要是集中在基礎的資料科學，包含進一步的資料呈現以及整理分析，之後隨著對資料科學技術的掌握，我開始將技術應用在我的研究中，而後我發現在資料科學領域中還有一門較為與我專業契合的領域稱為「文字分析」，而後隨著我對文字分析技術的熟稔，我漸漸將其擴大到語言處理，這裡就比較像是平時我們在提到 AI 時會舉的一些例子包含聊天機器人、文本生成等等。順著這條道路，相信大家也可以發現我並非一開始就選擇 AI 語言處理直接進入，而是循序漸進，從基本的整理資料開始培養起自己對於程式的掌握，而後慢慢延伸到較複雜的領域，但同時我也沒有包山包海的學習，在生活中我們一定聽過很多 AI 的例子，像是人臉辨識、偵測惡意回應，又或是這陣子如火如荼的 AI 作畫。AI

的應用範圍非常廣，AI 用更簡單的話來說，就是引導電腦來做人類能做的事情。而人類能做的事情非常多，小從單純的分類信件（例如分類垃圾信、廣告信），大至協助法官判決或是醫生診斷，因此 AI 並非離我們很遠，相反的，它應是能貼近我們的生活。因此掌握技術雖然重要，更重要的是知道要怎麼應用，而每個人需要的應用都不太一樣，我們的所學則為我們提供一扇窗，讓我們可以從自己擅長的領域切入生活，假設你的專長是輔導，或許你就可以開發一款 AI 他可以鼓勵他人以及根據不同情形提供適當的輔導資源。專業知識的結合才是 AI 價值的所在，因此如果你跟我一樣較晚才開始接觸這個技術，並不代表你過去時間浪費了或晚了人家一步，而是你花了比較多的時間去累積專業領域知識，這是我到後來才慢慢領悟的事情，而這樣的領悟也給了我很大的啟發跟動力，也希望這樣的想法能陪你度過程式前期的磨合。

我這次介紹的書雖然有些已經有中文翻譯，我非常建議直接閱讀英文原文書，由於這些比較像是手把手教學的工具書，因此裡頭的英文並不會太難。更重要的是，大部分我們在資料處理遇到的議題都已經在國外的程式論壇討論過，因此知道怎麼用英文敘述問題是非常重要的，哪怕只是幾個關鍵字，也是有希望能夠在網路找合適的解決方法，因此我們必須先知道每個步驟的英文是什麼，才能夠去敘述問題。

首先第一本我接觸的書是由 O'REILLY 出版社所出版的 *R for Data Science*，這本書的編

排是我以一個程式新手看來最舒服的，前期是以資料視覺化為主，也就會將一個像是 Excel 表格的實驗結果篩選然後畫出來，接著根據畫出的圖表去詮釋資料，這部分在前期是我成就感的來源，因為圖表出來之後，再來就是我們要透過專業知識去說故事了。在我們的研究中不乏問卷調查，或是翻閱或查閱大量資料後的觀察，如果我們能將這樣的資料透過一個更平易近人的圖表方式呈現出來，會比我們漫長的敘述還要來的強而有力，另外我覺得程式比 Excel 好的地方在於，他的圖表調整指令非常直覺簡單，只要按圖索驥，一定就能畫出你想要的圖，顏色形狀大小長度都只是程式裡的一小小行指令，對於資料呈現有執著的研究者而言，學會這套分析方式無疑是最大的福音，此外我們也能接住更多元的圖表類型除了傳統的點狀或柱狀圖之外，我們還可以遇見更多元的圖表類型像是點陣圖，它容許我們拿



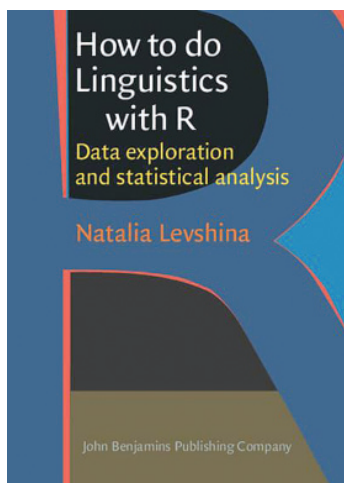
出三個變數出來比較以找出特定的規律。而歷經過繪圖的部分後才會進到複雜資料整理，同時也會慢慢串起資料整理跟繪圖之間的互動關係，慢慢跟著這本書用三個月的時間完成，我覺得就會對不同類型的資料以及呈現方式有一定的理解，如果手邊有正在進行的研究，也可以同時拿來試看看，會有更多啟發跟靈感！

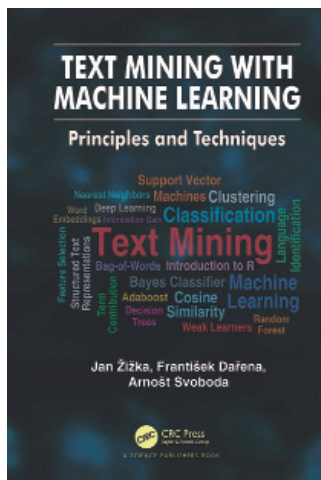
第二本我想推薦的書是 John Benjamins 出版社的 *How to Do Linguistics with R*，這本可以接續在上一本書後閱讀，這本書大多是語言學的研究且有很詳細的介紹不同類型研究可以用的分析方法，更重要的是這本書也同時教我們如何敘述我們的分析工具以及詮釋我們的資料。如前所說，擁有工具並不是唯一，更重要的是能說服讀者這是目前最合適的工具，以及有效的用自己的專業知識來詮釋結果。以往我在協助指導學生時看常常到一個很可惜的現象，學生用了一個

很棒的工具找出了有趣的結果，但他僅僅只是敘述實驗的結果而未加以討論，不免讓人覺得隔靴搔癢。而這本書每個章節都是一篇研究，我們可以找尋和自己研究類型最相關的研究，去看看這樣的工具如何被應用以及如何有效的解釋結果，而後以其為架構在有興趣議題上延伸。

接下來的兩本書難度較高，當時我自己一個人讀的時候常常一知半解，因此除了問老師或前輩之外，我還籌辦了讀書會。讀書會一方面能督促自己時時跟上團隊進度，另一方面也因為有組員分擔可以省下自己一直查找的時間，這兩本書分別是與機器學習還有深度學習相關，機器學習對於一個文組生而言是一個只聞其名不見其形的詞，感覺要有很厲害的程式基礎、很高深的數學概念，但其實一個更重要的概念便是，我們要知道怎麼用它，而知道怎麼用它並不代表我們需要完全理解知道其背後的建置原理，如同我們會開車並不代表我們要會組裝汽車。我們的目標應該是知道機器學習怎麼跟人文領域結合，而人文領域有可以怎麼在這一塊進行貢獻。

第一本是 CRC 出版社的 *Text Mining with Machine Learning*，裡頭淺顯易懂的介紹何謂演算法，演算法簡單來說就是一套教導電腦怎麼執行任務的方法，而這樣的方法在資訊科學裡頭是透過一個一個機率統計算式完成的，當然並不是強迫每個人要去理解每一個算式在做什麼，而是要去了解在哪个場合運用哪个方法才最有效。此外這本書的另一個焦點是 Text Mining，也就是文字探勘，任何有關文字的分析都可以被放入

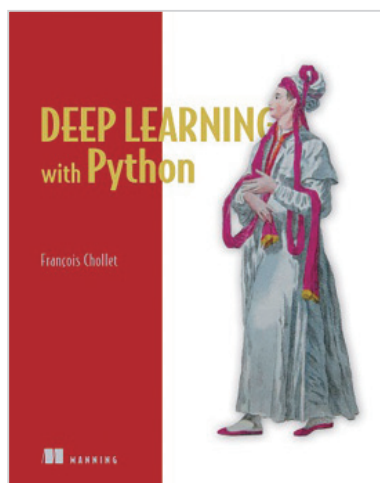




這個類別，例如偵測顧客文字表達後面的滿意度，他是滿意這個產品還是不滿呢？如果不滿的話是針對哪方面不滿呢？以往這類的任務我們需要聘請客服人員一則一則讀過並且歸類，但是透過一套有效的情緒偵測以及提取系統，我們就可以省下這樣的人力資源，因此在這時候我已經原則上確定了自己領域，因為我是語言學背景，因此我希望能更盡可能地將自己的程式學習導向跟文字相關的內容。接下來最後一本書，便是由 Manning 出版社出版的 *Deep Learning with Python*，Deep Learning 便是深度學習，其可以看程式機器學習的分支，只不過深度學習比我們之前看到的機器學習更方便的地方便是，在機器學習的過程中我們有時候看到機器誤差的部分便需要人為介入的調整，而深度學習則是機器裏頭會自己根據自己過往的錯誤而進行調整，因此到後期我們便會發現，我們只要鎖定特定的任務，

接下來便是決定用什麼方法找到答案而已，而在這篇文章介紹的四本書，分別是我在學習程式語言的四個階段的主要閱讀，但他們畢竟大多只是提供研究工具，而要用這樣的工具去探索什麼議題，其實很早之前就已經掌握在我們的手裡。

因此如果有讀者還在猶豫要不要修一門程式的課，我非常建議嘗試一番，抱持著以上的想法去面對程式，善用上述的額外閱讀，我們將會對自己的所學更有信心「原來這是可以結合的阿！」



延伸閱讀

1. Garrett Grolemund, and Hadley Wickham, *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data* (Sebastopol, CA: O'Reilly Media, 2017).
2. Natalia Levshina, *How to Do Linguistics with R: Data Exploration and Statistical Analysis* (Amsterdam: John Benjamins, 2015).
3. Jan Žižka, František Dařena, and Arnošt Svoboda, *Text Mining with Machine Learning* (Boca Raton, FL: CRC Press, 2021).
4. François Chollet, *Deep Learning with Python* (Shelter Island, NY: Manning, 2017).



秘密

アボガド 6 著

平裝本 / 11107/192 面 / 21 公分 / 380 元 / 平裝
ISBN 9786269604227/861

少年來到秘密圖書館，這裡的藏書收錄了各種人類與非人類捐贈的「秘密」。無法向人言說，卻也無法再承受更多的少年，也將自己的秘密留在這裡。他的秘密是什麼呢？噓，關於書的內容，還請保密……收錄 287 幅全彩畫作，主題包括人生光影、家庭悲喜、校園生活、醫病關係、自我認同，並收錄作者最受歡迎的動物插畫。致鬱也治癒，看似幻想卻又真實無比的 6 宇宙。看過各種秘密之後，最後揭露的少年秘密將讓你忍不住會心一笑！（平裝本）



雨衣娃娃 找雨鞋

管家琪 著

東方出版社 / 11104/96 面 / 21 公分 / 230 元 / 平裝
ISBN 9789863384786/863

大麗是一個身穿桃紅色雨衣和雨鞋的洋娃娃。有天小主人帶著她和家人一起去度假村裡玩。可是回到住處休息時，她發現自己的一隻雨鞋不見了！儘管玩具夥伴們不停勸阻，她仍決定獨自前往尋找雨鞋。走出小木屋，大麗發覺外面的世界非常陌生可怕，但她沒有放棄希望。忽然，大麗看到前方出現一個綠色小光點。大麗一路追著綠色的小光點，沿途遇到許多奇異可愛的動植物，雖然最後沒有找到雨鞋，但是雨鞋卻用別的方式回到大麗身邊。（東方出版社）

語言／文學



曙光 ：來自極東祕境的手札

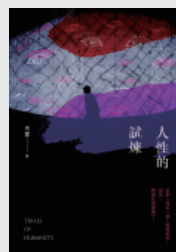
陳凱琳 著

三民 / 11105/272 面 / 21 公分 / 330 元 / 平裝
ISBN 9789571474397/863

這是臺灣第一本關於「馬崗」的文學作品。新銳作家陳凱琳，回到外婆家，透過紀實的手法，深入馬崗進行田野調查，提煉成一篇篇小說。這裡有乘風破浪的漁郎、堅毅美麗的海女，絮叨著土地與海洋，鹹甜的海風從馬崗吹來，挾帶海洋的訊息，告訴你馬崗的人、礁、風、雨。這本手札，記錄了馬崗「第一間」柑仔店的故事，保留了海龍傾的傳說，極力挽留馬崗人口老化後，逐漸消失的時空回憶。（三民）

語言／文學

語言／文學



人性的試煉

秀霖 著

要有光 / 11105/278 面 / 21 公分 / 340 元 / 平裝
ISBN 9786267058282/863

黑暗人性重重相疊的懸疑追殺令！峰迴路轉、局外有局！百變推理全才秀霖，耗費十五年打造，本土社會派犯罪小說最高峰！民風純樸的東部小鎮，一位早餐店老闆在報上新聞發現，不曾相認的女兒竟突然身亡。為了揭開她為何死亡之謎，單打獨鬥四處調查，卻遭到有心人士陷害，慘遭黑白兩道同步追緝。醫療糾紛與政治惡鬥更牽扯出令人意外的一連串陰謀……穿梭奔走之中，平民百姓有辦法與黑手無處不在的權貴抗衡嗎？（要有光）

語言／文學