

教育部教學實踐研究計畫成果報告(封面)

Project Report for MOE Teaching Practice Research Program (Cover Page)

計畫編號/Project Number：PSL1080052

學門專案分類/Division：社會(含法政)

執行期間/Funding Period：108/08/01~109/07/31

當統計遇見文學:培養人文系所學生具備大數據知能之教學實踐行動研究計畫

An action research project of teaching practice for cultivating students in the Department of Humanities with the “Big Data” knowledge and capacity

配合課程名稱:1.文本探勘在決策上之應用

2.文本探勘小技術大應用」摩課師課程

計畫主持人(Principal Investigator):許健將

共同主持人(Co-Principal Investigator)：

執行機構及系所(Institution/Department/Program)：

中興大學教師專業發展研究所

成果報告公開日期：

立即公開 延後公開(統一於2022年9月30日公開)

繳交報告日期(Report Submission Date)：109.09.15

當統計遇見文學:培養人文系所學生具備大數據知能之教學實踐行動研究計畫

An action research project of teaching practice for cultivating students in the Department of Humanities with the “Big Data” knowledge and capacity

一、報告內文:

(一) 研究動機與目的:

本研究的發想主要來自於以下幾項因素:首先從歷年選修本課程的學生人數的改變情況來進行探討，課程的修課人數自 105 學年度開始的增加，至本 107 學年度已達 18 名之多(註:本所每年招生 13 名學生)，成長之幅度相當可觀。主持人仔細反思造成這種改變的一項很重要原因是:

自 105 學年起，主持人就開始改以 R 語言為工具，進行本課程的教授，或許就如 105 學年度的學生意見回饋中所述:『在 R 語言的摸索中，能發現到不同指令下所能呈現的許多想像不到的功能，是一個似乎沒有極限且直觀的開放性功能語言。』而能夠得到學習者的青睞。此外，在教學過程中發現，由於 R 語言所具備的強大繪圖功能，更是經常能吸引學生的目光，應該也是導致有愈來愈多學生願意選修此門課的重要因素之一。然而，儘管修習本課程的學生人數成數倍的成長，但在修習本課程後，真正利用 R 語言為工具進行相關量化研究的學生卻仍屬極其少數。探究其原因發現，每年進入本所碩士班的學生以外文系、中文系和教育相關系所之大學畢業生為主，因此，對於需使用到統計相關技術的研究，總是先入為主的排斥而顯得興趣缺缺。主持人為了破解這些文科背景學生腦海中的迷思，即文科生不擅長或不適宜進行與統計相關的量化研究，擬透過本教學實踐研究計畫，利用過去三年已有的 R 語言教學的經驗(已證實 R 語言對文科背景的學生具有親和力)，以目前大數據時代之「文本探勘」

技術的培養，為本課程設計的主軸，期待透過本研究的成果，對如何培養文科背景學生跨領域能力，增加其未來競爭力提供一項新的思考方向。

(二) 文獻探討

文本探勘(Text Mining)的基本定義為:從大量的文本數據中，透過統計技術，對其中隱藏的、過去未知的以及潛在有用的資訊進行有意義的萃取(Waegel, 2006)。文本探勘是一項新興的研究技術，涉及各種數學，統計，語言和模式識別技術的發展。它試圖從自然語言文本(Natural Language Text)中提取有意義的資訊。與傳統存儲在數據庫中的數據類型相比較，文本(text)是屬於非結構化、模糊並且較難處理。然而，在現代文化脈絡中，文本卻是訊息傳遞中最常見的使用的方式。由於進行文本探勘所涉及的對象通常是真實的文件或意見交流的文本，試圖從這樣的文本中去自動提取資訊的過程本身就相當令人著迷(Kumar & Bhatia, 2013)。

文本探勘類似於數據探勘(Data Mining)，然而數據探勘工具只處理來自數據庫中的結構化數據(Structured Data)，但文本探勘還可以處理非結構化(Unstructured)或半結構化(Semi-structured)數據集，如電子郵件(e-mail)，文本文檔(Text Documents)和 HTML 文件等。因此，文本探勘相較於數據探勘是一項更強而有力的分析技術(Navathe & Elmasri, 2000)。文本探勘通常是對輸入的文本進行結構化的過程，然後在結構化數據中導出模式，以及對輸出的結果進行評估和解釋。

(三) 研究問題

1. 如何有效促進人文、社會與法政背景學生程式語言的學習效能？
2. 利用「文本探勘」技術為手段之教學設計，是培養人文、社會與法政背景學生程式語言設計能力及認識大數據應用有效策略？

(四) 研究設計與方法

本研究《當統計遇上文學》，係以建構理論之「過程導向引導探究學習」(Process Oriented Guided Inquiry Learning)為基礎，以 Coghlan 和 Brannick(2005)所發展之後設探究循環(Meta cycle of inquiry)行動研究模型為方法，以互動式講授(interactive lecture)與反思(reflection)與為策略，融入半學期十八周之「統計套裝軟體應用」課程，藉以培養本校人文、社會與法政相關系所碩博士生運用 R 語言之「文字探勘」(text mining)技術進行相關小說、著作或文件之分析能力，促進其對「大數據」在各該領域未來之發展與應用有更深層之認識，促發其對培養跨領域能力的重視，增加其就業競爭力。為達此目的，將採示範、指導、提供鷹架、闡明、反思與探究等流程進行本課程之教學設計，並輔以撰寫反思周誌、深度訪談等方法，以及對整體研究過程之內在與外在評鑑，藉以探討學生參與本教學實踐研究計畫後，其運用 R 語言能力與文本探勘技術之成長機制，以及主持人對此教學實踐研究計畫之省思。

(五) 教學暨研究成果

1. 教學過程與成果

- (1) 由於 R 語言的語法直觀、沒有複雜的數學符號與運算且相關套件豐富多元，非常適合人文與社會相關系所背景之學生學習。
- (2). R 語言可免費下載使用，非常適合尚未具備經濟基礎的學生使用。
- (3).相關討論 R 語言的網路資源或部落格相當多，學生在學習過程中若遇到困難，在課後也可以很快得到協助解決，有助於學習成效。

2. 教師教學反思

- (1).透過本研究計畫，個人所編製完成之《文本探勘小技術大應用》磨課師課程，獲中興大學 108 學年度校內磨課師課程設計優良獎項。
- (2) 透過本研究計畫，個人所編寫完成之《文本探勘小技術大應用》專書，亦將於 109 年 9 月份由全華圖書公司出版發行。

3. 學生學習回饋

- (1). 學生普遍認為本課程資料豐富完整。
- (2). 由於學生先備知識差異性頗大，在學習上的成效有明顯落差，未來在教學設計上因思考如何進行調整。

(六) 建議與省思

由於教學是持續精進的過程，建議本教學實踐研究計畫能盡早開放多年期計畫之申請。

二、參考文獻

- Coghlan, D., & Brannick, T. (2005). *Doing Action Research in Your Own Organization* (Second Edition Ed.). London: Sage Publications.
- Kumar, L. & Bhatia, P.K. (2013). Text Mining: Concepts, Process and Application. *Journal of Global Research in Computer Science*, 4(3), 36-39.
- Navathe, S. B & Elmasri, R. (2000). *Data Warehousing and Data Mining*, in *Fundamentals of Database Systems*. Pearson Education Inc, Singapore, 841-872.
- Waegel, D. (2006). *The Development of Text-Mining Tools and Algorithms*. PA: Ursinus College Press.

三、附件

與本研究計畫相關之研究成果資料，可補充於附件，如學生評量工具、訪談問題等等。

