

運用文字探勘技術於探索護理情境教學 師生互動行為議題與發展

趙莉芬（通訊作者）

長庚科技大護理學院臨床技能中心、護理系副教授

暨林口長庚醫院急診醫學部副研究員

E-mail: lfchao@mail.cgust.edu.tw

陳俊良

國立臺北科技大學運動健康休閒研發中心顧問

E-mail: kennyqoo@gmail.com

黃湘萍

長庚科技大學護理學系副教授兼護理科主任暨課程組副主任

E-mail: hphuang@mail.cgust.edu.tw

倪麗芬

長庚科技大學護理學系副教授

E-mail: lfni@mail.cgust.edu.tw

曾瑛容

長庚科技大學護理學系助理教授

E-mail: yjtseng@mail.cgust.edu.tw

摘要

課室教學師生互動行為歷程分類與定義，多是以專家觀點設定質性對應類別，然分析運用設定的類別實容易發生無法解釋，且缺少量性訊息提供教學應用的具體建議。本研究目的以運用文字探勘技術，並採質量混合研究設計多重檢核方式形成的研究問題，進行蒐集資料、分析資料與詮釋結果。將蒐集後的資料，透過數據向量化後所計算的資訊熵值，發展出一套研究教學輔助指引與

師生互動行為記錄分析資訊化系統，以探索關懷、溝通、邏輯思維等專業核心素養，於教學師生互動行為的特徵並建立常模。執行過程中已建立之斷詞詞庫，包含 196,392 詞彙與 99,231 符號，擬將加入學術電子資料庫與碩博士論文等線上檢索檔案文章，進行文字探勘作業，以作為課室教學溝通互動類別設定之依據，使發展出的系統更具信效度。

關鍵詞：質量混合研究設計、資訊熵值、文字探勘

壹、前言

教學是在教育目標下，一種有計畫、有目的的師生互動歷程。教學是老師與學生間行為模式與價值觀念互動的歷程，其中雙方的語言溝通是影響教學效果的關鍵因素（Worku & Alemu, 2020; 簡馨瑩, 2010）。在任何課程領域的學習歷程，師生合宜具深度的互動，使得教與學之間能產生有效的連結，成為老師教學的動力來源，學生因而能獲得充分的學習。近年來，醫護教育採用互動溝通教學方式的臨床擬真情境學習（scenario-based learning）取代傳統教學型態，已成為世界先進國家改革醫護教學的主流模式（Chao, Huang, Ni, Tsai, & Huang, 2017），以學生為中心的情境教學是依據學習目標，依循學習者的先備知識，設計以臨床問題解決策略導向的案例教案，教師安排適當的教學環境與情境結構，讓學生身歷其境實務進行臨床推理、互動式情境操演與回饋反思活動，建構學生臨床實務核心能力。而回饋反思過程階段，護理教師與學生互動的歷程，雙方的語言溝通，被視為達成情境教學目標的主要關鍵。技職教育教學者能熟稔掌握專業技能與專業情境，然而以親身示教的身教外，如何於課室中以言語表述或傳達專業知能，可讓互動溝通引導出有效的學習刺激，是教師冀待強化的議題。

近年來對電子化訊息解讀的需求陡升，促使資料探勘（Data Mining）的處理，進一步延伸出文字探勘（Text Mining）的應用。文字探勘（Text Mining）是隨著人工智慧和自然語言處理技術發展的一門新興技術。為蒐集的特定巨量文件，執行編輯、組織與分析的過程，以發現其間隱含的特徵關聯或新穎有趣的模式，提供主題應用人員或決策者特定的訊息。主要從大量、模糊的、隨機的文字資料中自動化辨識與挖掘有用的資訊，萃取出隱含的、規律性的、人們事先未知的，潛在有用的，且最終可以理解、可信與有效的訊息。資料分析是透過統計與數學技術架構來輔助，不需要先建立假設，並且依據使用者文字表達特徵，在一群未經處理的資料中找到使用者可能感興趣、過去未知和不可預測的結果（Hang, Liu, Ren, & Dai, 2005; 張云濤、龔玲, 2007; 黃仁鵬、張貞瑩, 2014）。文字探勘和一般用統計方法來進行資料分析在運用目的上是不同的，統計分析在於要用單一規則或模式顯示資料的性質，但大多數的資料是由多條規則或多個模型混合

產生的。文字探勘則是除了運用統計的概念來判讀資料的性質外，更強化用電腦的高運算效能，來歸納哪一條規則或模型適合用來解讀其一部分的資料，如此可使資料的詮釋更具彈性，也較能發掘出真正隱藏在資料背後的資訊，是以統計學和電腦科學為基礎，所發展出來能快速分析資料的方法（翁慈宗，2009），運用的資訊對解決問題或規劃未來的發展上有相當大的參考價值，近年許多學者也應用文字探勘技術至各種社會科學研究（丁怡婷、劉志光，2010；張曉珍、李永銘，2012；梁衍忠，2012；童瓊慧，2008）。

合宜的師生互動模式透過系統觀察法具體的、完整的記錄與觀察，提供的不僅是師生之間的互動模式，也是給予教師在教學技巧、課室經營、教學內容及教學模式，提供一個具體量化的數據呈現，相信對於教師的教學成效與品質必有莫大的幫助。本研究主要目的是運用文字探勘的技術，探索分析模擬情境課室教學中師生互動行為歷程分類與定義，測試與建立師生互動行為常模與建議模式指標，未來可透過建議模式指標，建立具體化教學運用策略參考指引，且研發出自動資訊化的文字紀錄分類分析系統，以提供做為教學評值、教師培訓與教學研究之運用。



貳、文獻探討

CACET

中華資訊與科技教育學會

一、高擬真護理情境教學

傳統護理教學多使用部分人體模型練習器，或要求學生以假想情境於同儕間角色扮演，模擬護病溝通，這種方式較適合學習簡單技能或操作特定流程，卻難培養學生全面性獨立思考，成為具有解決問題能力之專業人才。教育改革所提出的「情境認知」論述，強調模擬真實情境的體驗學習（experiential learning），能幫助學生建立完整的知識體系，不僅提高認知程度，並能促進專業判斷、知識應用及整合性能力（Johnson, 2020）。醫學擬真（medical simulation）是指運用機器人或模擬電子人於臨床醫學實習，美國國家醫學研究院（Institute of Medicine, IOM）於1999年報告即建議盡量使用情境模擬（simulation）進行教育訓練，以兼顧病人的安全與醫療品質的提升（Sharma, Boet, Kitto, & Reeves, 2011）。擬真體驗學習是以學生為中心的教學方式，讓學生在不影響真實病患安全的時空裡進行探索，由學生主導病人照護，進而發展出多種問題解決策略，並可以透過重複練習增加技巧的熟練度（Brigden & Dangerfield, 2008; Kneebone, Nestel, Vincent, & Darzi, 2007）。國內醫護教育相繼導入此教學策略與教學環境，筆者任職之教學單位近年建置的擬真護理情境包含急重症加護單位、內外科、老人居家、產科與兒科等各科模擬情境教室模組。各組皆配置有該科高擬真電子模擬人教具，並

輔以臨床環境的模擬來進行教學。依情境處置進行應對轉換反應，幫助學生在符合真實的情況和脈絡下學習。能實際操作評估、測量與持續監控的臨床實務，就像面對真正的臨床病人，在不斷與情境互動的過程中，建構出自己的專業知能與態度（Chao et al., 2017; 黃湘萍、趙莉芬，2018）。在教學環境空間規劃上，分為擬真病房單位與學生課室教學區，讓在校學生能感知與熟悉病房作業單位，執行專業人員角色功能，於一般教室則授課教師示範技能與情境演練後回饋互動。護理教師則需於課前發展適當的教案，安排情境相關的角色與環境；熟悉情境教學策略，靈活操作數位教具，引領學生將理論應用於擬真臨床情境的實作中，培養學生於傳統教學較缺乏之溝通與團隊合作、批判性思考與反思能力。

二、資料探勘技術之概念與相關理論

一般監督式學習為將手中資料區分為三個獨立的部分，利用訓練資料（training data）作為機器訓練用，再以受測資料（testing data）作為測試評估與建模之用，最後再對未經標記的目標資料（target data）進行預測（陳世榮，2015）。若以驗證評估為主，資料僅區分訓練與受測資料，由於檢索出的資料屬於自然語言，傳統內容分析法則先對所有資料進行人工分類標記，標記結果可作為訓練資料，以提供機器學習之用，且人工標記結果納為受測資料時，將可提供做為文字探勘與內容分析的一致性檢驗。資料筆數愈大，則較能確保文件分類的效果與模型的精確性（Witten, Frank, Hall, & Pal, 2016）；但現實狀況上則是總會面臨資料量少，納入分析的資料量通常未必是大的。

採用時間區段檢索出較小量的文件集，藉此壓縮機器學習的空間以瞭解文字探勘在小文件集的分析效果，可藉以觀察監督式分類器的穩定度與應用廣度。在效果評估後，再透過所見問題做進一步分析推論，可從而提出相關的注意事項與建議，是較為實際的應變處理，本研究即採此方式進行。要能從資料中挖掘出有用且非顯而易見的資訊，探勘的進行主要分 5 個步驟。第 1 步為目標設定：確定資料探勘的目的，確認以計算或資料檢索都無法產生所需要的資訊後，才適合使用資料探勘；接續進入第 2 步驟即蒐集不同來源和格式的資料，整理出適合探勘的資料：資料的來源有很多種，包括資料庫中的檔案或大量歷史資料，或是網頁資料；資料型態除了傳統的文字與數字資料外，還有網頁格式、語音和圖像資料。第 3 個步驟是進行資料的前置處理：蒐集好的資料並不見得就可以進行資料探勘，因為資料中可能有些有異常，或是有些欄位的值無法取得，這些都會直接影響到資料探勘的結果。因此。這個步驟除了處理異常值和遺漏值外，還要考量欄位是擺放數字還是文字，以及每個欄位的必要性。另外為了避免欄位的度量對結果造成影響，還要考慮是否進行資料的正規化。例如衡量長度時使用公尺或公分，會

使該欄位有一百倍的差距，而這一百倍的差距有可能嚴重影響到該欄位在資料探勘時的重要性。欄位轉換則是考量到有時候必須結合多個欄位才能產生有用的資訊。例如單純只用年齡、身高或體重單一欄位，很難判斷一個人是否過重，但如果把這三個欄位轉換成體脂肪一個欄位，這個新的欄位會更適用來判斷一個人是否過重。因此若能做出適當的欄位轉換，會更容易取得有意義的探勘結果。對於一開始設定的目標，要了解適用的資料探勘類型是哪一種，否則無法得到有用的資訊。甚至往往為了能得到有意義的探勘結果，必須結合多種類型的資料探勘工具。因此了解各類型資料探勘的定義和資料分析方法後，才能選擇適當的工具進行資料探勘。最後對於探勘結果的詮釋，則要倚賴分析人員具備的背景知識來解讀探勘的結果。此部分須由資料探勘的人員和產生資料的人員合作，除了能適切地詮釋探勘結果外，也才能知道該如何把結果應用在實務上。

若發覺探勘中的某一步驟有問題，則應考慮回到上一步驟去檢討和修正。例如認為資料探勘的結果不理想，可以考慮使用不同的正規化方法來處理資料或進行欄位轉換，這樣或許可以產生較佳的探勘結果。如果還是不太理想，可能是因為有些有用的欄位沒有被選取，如果能回到資料取得的步驟重新檢視，或許所得的結果就能讓人耳目一新。幾乎在所有的資料探勘案例中，都要不斷地回到上一步驟去檢視，並進一步從探勘結果中思考應如何修正才能獲得較佳的結果。在如此不斷來回的測試中，累積對資料探勘的經驗，也才能得到滿足當初所設定目標的結果。

中華資訊與科技教育學會

三、教室內互動要素設計

師生是教室內的學習主體，教師與學生的口語溝通則被認為是達成學習目標的主要方法之一。讓學生能有效地系統化、內化課程教師的語言資訊，而能在課室過程中有說話和表達想法的機會，營造了創造思考的空間，進而能夠在真實的社會情境中，展現所學的行為互動模式，與他人作資訊、觀點、情感等之交流，也產生知識分享的作用。教室內教學的主要是教師所提供的資訊，能有效的傳輸給學生，師生間的溝通對話，隱含著教師的教學型態、學生學習方式、彼此的互動關係與知識的建構，被認為是潛在但為豐富的研究素材（陳昇飛，2013）。

分析教室內言談結構的主要目的，是觀察教師如何透過與學生溝通互動的一連串言談，來進行教學活動，並且如何將這些言談結構串成一個完整的教學主體，以求達到所預期的教學目的。教學基本上，是由教室內的每一個言談動作串接起來的，觀察一個有效的教學，可以從檢視它的活動流程開始，基於每一個言談舉動都會有其特定的目的，因而檢視這一連串具有教學目的和意義的言談動作，藉以瞭解其教學目標如何融入實際教學的全貌。

四、數值運算模式採用概念

有關本計畫所設計之建議模式指標設計的運算，試著運用 information entropy 的計算概念（Ben-Naim, 2017），將收集的資料完成文字探勘後，建立其多維的向量數值，並利用 information entropy 演算後，以資料離散程度來定義建議模式指標。

在資訊理論中的熵（entropy）是指接收的每條訊息中包含的資訊的平均量，又被稱為資訊熵、信源熵、平均資訊本體量。「訊息」代表來自分布或數據流中的事件、樣本或特徵，熵最好理解為不確定性的量度而不是確定性的量度，因為越隨機的信源的熵越大。來自信源的另一個特徵是樣本的機率分布（Liang, Shi, Li, & Wierman, 2006）。這裡的想法是，比較不可能發生的事情，當它發生了，會提供更多的資訊。由於一些其他的原因，把資訊（熵）定義為機率分布的對數的相反數是有道理的。事件的機率分布和每次事件的資訊量構成了一個隨機變量，這個隨機變量的均值（即期望）就是這個分布產生的資訊量的平均值（即熵）。熵的單位通常為位元，但也用 Sh、nat、Hart 計量，取決於定義用到對數的底。

依據 Boltzmann's H-theorem，把隨機變量 X 的熵值 H。定義如下，其值域為 $\{x_1, \dots, x_n\}$ ：

$$H(X) = E[I(X)] = E[-\ln(P(X))]$$

其中，P 為 X 的機率質量函數（probability mass function），E 為期望函數，而 I(X) 是 X 的資訊量（又稱為資訊本體）。I(X) 本身是個隨機變數。

當取自有限的樣本時，熵的公式可以表示為：

$$H(X) = \sum_i P(x_i) I(x_i) = - \sum_i P(x_i) \log_b P(x_i),$$

在這裏 b 是對數所使用的底，通常是 2，自然常數 e，或是 10。當 $b = 2$ ，熵的單位是 bit；當 $b = e$ ，熵的單位是 nat；而當 $b = 10$ ，熵的單位是 Hart。

當 $p_i = 0$ 時，對於一些 i 值，對應的被加數 $0 \log_b 0$ 的值將會是 0，這與極限一致。

$$\lim_{p \rightarrow 0^+} p \log p = 0$$

還可以定義事件 X 與 Y 分別取 x_i 和 y_j 時的條件熵為

$$H(X|Y) = - \sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(y_j)}$$

其中 $p(x_i, y_j)$ 為 $X = x_i$ 且 $Y = y_j$ 時的機率。這個量應當理解為 Y 的值前提下，隨機變量 X 的隨機性的量。

「資訊理論」中的熵，是資訊壓縮理論的基礎，不同於熱力學中的熵。熱力學中的熵是描述系統中不能做功的能量總和，亦做為描述系統中失序、不穩定的指標；熵則是描述事件（系統）的訊息量，其代表著事件各種結果的可能性，當事件越明確，則代表其不確定性越小、事件的結果越穩定，訊息越少，熵值就越小，反之熵值則越大。

參、研究方法

擬以質量混合研究設計（mixed-methods research）多重檢核方式，針對所形成的研究問題，進行蒐集資料、分析資料或詮釋結果。以監督式機器學習對非結構、複雜文本的分類效果，以課室教學師生互動特定議題為對象，遵循文字探勘一般程序，驗證其文件意義分類的效果，並發現可能問題，以提供未來比較與改進的經驗依據。探索關懷、溝通、邏輯思維等專業核心素養在教學師生互動行為的特徵並建立常模，並就所見議題提出更進一步的實證比較分析與建議，更進而發展互動行為分析資訊系統。

斷詞詞庫與互動教學歷程影像資料收集

斷詞詞庫資料收集上，現階段是以網路新聞為主，透過現有網路新聞上的分類類別進行相關詞彙的收集與分析，作為斷詞運用上的詞庫依據。為了讓整體詞庫範圍與信度更廣泛，下階段將預計利用學術電子資料庫與碩博士論文等線上檢索檔案所提供的課室教學溝通互動事件進行文件選取與文字探勘的來源文章。

課室互動教學歷程影像錄製部分，則是以方便取樣方式，主要選取筆者學校護理系教師與其上課班級的課程事件，回溯使用蒐集情境教學回饋互動過程的課堂影片分析為資料來源，針對課堂觀察教師與學生的語言互動行為呈現的教學樣態，研究的對象包括 2018 至 2019 年接受情境教學的學生與教師，科目為護理專業課程科目，採現場錄影而後將語言互動過程轉成文字記錄稿，並撰寫非語言情境過程互動資料，以便於多角度瞭解情境，增加詮釋的正確性與意義的再現。在課室教學過程轉成文字稿後，由研究者先比對資料內容，找出每次溝通互動主

題，再依據每次師生互動找出呈現對話中的關懷現象；初步分析結果交由同事及研究對象審閱，之後再澄清不清楚的項目與不明確語句的書寫，並在討論中說明，增加研究結果的可靠度。

肆、結果與討論

一、斷詞詞庫現階段成果

本計畫執行期間，透過網路爬蟲程式於網路上進行大量文章的抓取，再將文章標題與連結合併成 data frame，透過 TFIDF 運算（term frequency-inverse document frequency，是一種用於資訊檢索與文字挖掘的常用加權技術，簡稱 TFIDF）分析收集相關資料後，載入斷詞分析的套件（jieba），建立起自有的斷詞程式與詞庫。目前透過計畫收集的詞庫總共有 196,392 詞彙與 99,231 一字詞與符號。四大類別的 data frame（包含 term_id、domin df、tdomin、newweight、name 等屬性欄位）。有關斷詞分析的套件（jieba）所建立的雲端系統如圖 1。這是一套雲端程式，使用者只需要透過網路連線至此系統，即可進行自動斷詞、定義詞、批次作業等程序。

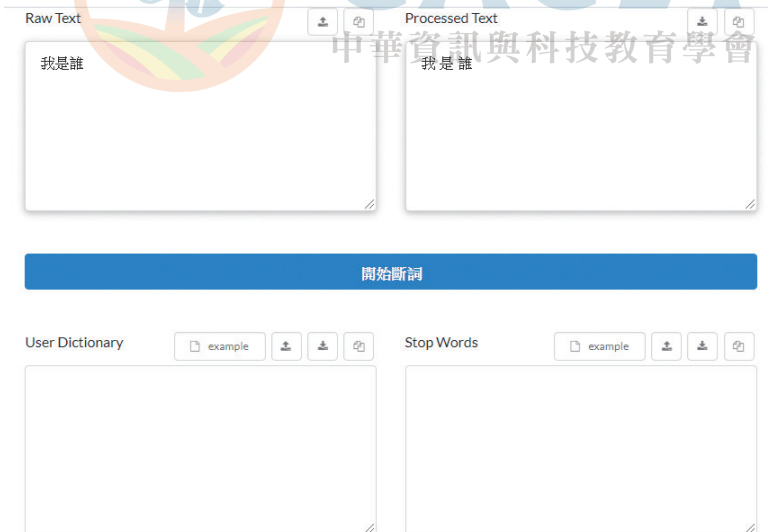


圖 1 詞意定義類別

未來這些詞庫的運用，除了使用於斷詞程式上外，也可作為下個階段語意分析時的權重調整依據。

二、教學歷程影像資料成果

將情境教學 Debriefing 回饋階段之師生互動內容謄寫成逐字稿後，研究者根據是運用本文資料脈絡的區分類別屬性，對所收集資料加以分類反覆仔細閱讀逐字稿，找出與課堂中師生互動且有意義的陳述，分類編碼與分析步驟進行資料分析資料飽和，逐步形成發展脈絡。綜合整理各科情境課程資訊，以逐字稿編碼後，以專家會議進行歸類教師話語類別運用可分為 28 項概念，5 個類別與 2 大面向。

三、建議模式指標設計的可行性運算說明

資料類別向量數值設計上，於本計畫中是採用三層架構基礎設計，如下圖 2。

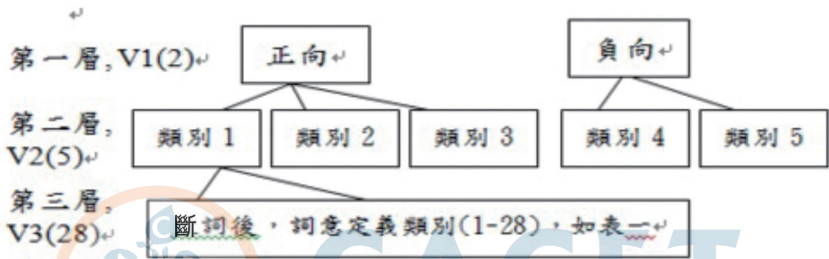


圖 2 建議模式指標設計詞架構

所以每筆收集的教學歷程影像資料對應之教室內互動要素，就會有 3 個基本的向量數值。另外，本計畫重點在於探討互動溝通評值，因此也特別將老師與學生互動過程頻率，定義成 V4(2) 的向量參數。各階層向量頻率定義為：

- V1(2)=(正向、負向)
- V2(5)=(類別 2(共 5 類))
- V3(28)=(類別 3(共 28 類))
- V4(2)=(教師互動 , 學生互動)

譬如：以某一反思討論課程而言，老師與學生互動過程頻率如表二。利用表二將特徵值應用於這一個反思討論課程而言，

- V1(2)=(21,1)
- V2(5)=(14,4,3,1,0)
- V3(28)=(11,0,0,0,1,0,0,3,1,3,0,0,0,0,1,0,0,0,0,1)
- V4(2)=(11,11)。

將向量標準化後，直落於 0~1 之間，且向量總和必須為 1。

- V1(2)=(0.5,0.5)、
- V2(5)=(0.64,0.18,0.14,0.04,0)、

$V3(28)=(0.5,0,0,0,0.04,0,0,0.15,0.04,0.15,0,0,0,0,0.04,0,0,0,0,0.04)$

$V4(2)=(0.5,0.5)$ 。

當有多組資料一起計算轉換下，就會產生四組 V1、V2、V3、V4，並各自計算出事件相關的比例係數，與後項的事件發生機率，如下公式

$$H(X|Y) = - \sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(y_j)}$$

表 1 詞意定義類別

第三層 (詞意類別)	第一層	第二層
1.發問、問問題 (a分類:整體直接詢問? b比較式問? c提示、導向式詢問)	正向	類別 1
2. 給予指示或指導	正向	類別 1
3. 關注 focus(整體 a、具體 b)	正向	類別 2
4. 反應	正向	類別 2
5. 提供課堂資訊 (介紹情境環境 b)	正向	類別 2
6. 客套，段落結語	正向	類別 2
7. 接納	正向	類別 2
8. 自我澄清	正向	類別 1
9. 尋求確認	正向	類別 1
10. 給予表揚 (整體)	正向	類別 3
11. 轉換	正向	類別 3
12. 累加挑戰	正向	類別 3
13. 具體的讚許	正向	類別 3
14. 停頓靜止 (疑惑、整理思緒)	負向	類別 4
15. 教師表示參與其中、關心 (具體)	正向	類別 2
16. 鼓勵誘發	正向	類別 3
17. 幽默 (自嘲)	正向	類別 3
18. 批評、批判	負向	類別 4
19. 講演 (內容說明 19、依對方反應延伸說明、解釋 19b、自問自答 19c)	正向	類別 1

20. 權威否定	負向	類別 4
21. 期許	正向	類別 2
22. 表達理解、認同	正向	類別 2
23. 不同意、反對、不認同	負向	類別 5
24. 忽略對方反應	無反應，講自己的；不停留、給對方時間說明（負向）	類別 5
25. 合理化解釋	正向	類別 2
26. 提供選擇機會	正向	類別 2
27. 帶動、引發全體氣氛	正向	類別 1
28. 提醒	正向	類別 1

表 2 反思討論課程師 (Tr) 生 (St) 互動頻率

角色	類別	原句用語
St	講演 (內容說明)*	就是有統計說 cad 的這群病人
Tr	發問、問問題 *	請問有同學要提問嗎
St	發問、問問題 (比較式問)*	那如果沒有第一期呢?
St	肯定明確的說明 *	一定要達到 100% 這樣子
St	發問、問問題 (提示、導向式詢問)*	這個部分可以再說明嗎
St	具體的讚許 *	學姊的台風很穩健
St	具體的讚許 *	口條也非常的清楚
St	具體的讚許 *	時間控制得也非常地好
Tr	發問、問問題 (整體直接詢問)*	我想釐清一下你的 y 面項是什麼，怎麼接
Tr	尋求確認 *	完整參與
Tr	發問、問問題 (比較式問)*	如果拆掉架一半呢
Tr	發問、問問題 (提示、導向式詢問)*	請你思考
Tr	教師表示參與其中、關心 (具體、)	我只好奇
Tr	關注 *	我們其實在探討影響因素的時候
Tr	自我澄清 *	常常要探討可改變的變項關係
Tr	發問、問問題 (提示、導向式詢問)*	這個在哪裡

Tr	肯定明確的說明 *	所以那一個量表其實也要叫做醫療可近性的量表
St	發問、問問題 (整體直接詢問)*	你是用家訪的吧
Tr	發問、問問題 (整體直接詢問)*	一定會回來那個醫師這邊
St	發問、問問題 *	萬一在新竹呢
St	發問、問問題 (整體直接詢問)*	還要回來啊
St	權威否定	太麻煩了

當收集的資料透過轉換運算後的 **information entropy** 熵值後，再下個階段也可以透過以下統計方式進行類別群組的定義。其方式分別為群集分析與區別分析。各自演算方式描述如下：

• 群集分析

群集分析主要應用在多變量資料上，為多變量分析中相當實用的分析工具之一，其目的為發現或偵測資料中的群聚現象，使得每一集群中資料有高度的同質性，而不同集群中有較大差異。集群分析已經被廣泛應用在醫學、自然科學、基因體學及經濟學上的資料分析，是分析高維度資料及大型資料庫的重要資料探勘 (data mining) 工具之一。

下階段導入進行兩大步驟：（一）相似量度 (proximity measure) 的選擇：首先要根據分析的目的來定義衡量任兩筆資料相似度的測量值，相似度越高表示資料越有可能屬於同一族群。（二）群集演算法的選擇：根據相似量度，群集演算法會將資料做分群，使得群內差異小，而各族群間差異大。

• 區別分析

線性區別分析 (Linear Discriminant Analysis, LDA) 是一種可用來做分類 (classification) 的 supervised learning 方法。這個方法主要用來區分資料中群體的差異，當分析的資料中包含依變數與自變數，且依變數為類別 (categorical) 型態，而自變數為數值 (numerical) 型態時適用此方法，方法中選擇合適的自變數為判別變數並建立判別函數 (discriminant function)，再使用此函數資料加以分類。區別分析是多元統計分析中用於判別變數所屬類型的一種方法，與集群分析相同的勢將相似的樣本歸為一族群，不同處卻在於集群分析預先不知道分類，而區別分析是在研究對象分類已知的情況下，根據樣本資料推導出一個或一組區別 (判別) 函數，同時指定一種判別規則，用於確定待判別樣本的所屬類別，使錯判率最小。

針對下階段執行上的建議，除了持續收集更多教學歷程影像資料外，就是 **information entropy** 計算後的數值應用。搭配上述統計分析的方法加上語意分

析與主題類別的設定，進行計畫過程中訓練資料（Training Data）與測試資料（Testing Data）的分類判析，最後可產出研究教學輔助指引與師生互動行為紀錄分析資訊化系統，並希望透過更多資料收集與應用建立師生互動行為常模，建立具體化教學運用策略參考指引。

目前教學資料互動形式，可分析出課室互動的類別，多因教師個人因素，而在互動行為類別上有區別的趨勢，不同課程科目在情境回饋上互動時所應用之語意脈絡，無特別差異。師生互動在社會情感交流部分的溝通比率上較低，沒有超過 15%，多以任務導向的溝通居多；另正向類別較多超過 80%，負向語意脈絡則較少。未來在實務教學時，需要強化護理學的同理心和體察人心的能力，將有助於護理學生持續保有以老人為中心的溝通能力，並且可能可以提升護理學生在溝通過程多和個案有情感性交流的機會。

本計畫及依原定行程進行資料分析與界定類別，依規劃完成質性教學專家學者焦點團體分析討論法擬定探勘的議題目標，並檢核調整探勘訓練資料所得之初始模型。將繼續研討印證比對機器學習之量化分析的結果，以綜合質化與量化資料，調整出最通常模，作為教學上可應用的具體化建議。過程中，研究者除實踐研究方法學以外，更是初始在履行 AI 的應用，目前進行階段為資料的深度學習。預期將於近期完成初探過程結果的文章，以跨領域合作形式取向做期刊雜誌目標；而研究成果是以人工智慧方式提供課室教學師生互動增進學習成效的相關知能，可發展出行為紀錄分析資訊化系統，將成為提供教學應用研究的創新工具，是本計畫可預期的價值。

中華資訊與科技教育學會

本研究首創將巨量資料文字探勘此新興技術，運用於專業核心教育跨領域研究應用。主持人及團隊整理相關結果，以發表創新教學學術論文期刊與參與國內外學術研討會，促進教學應用學術交流。參與人員藉此獲得課室教學師生互動增進學習成效的相關知能，並學習文字探勘研究工具技巧、訓練，以及高級統計數據分析統整模式與轉換成記錄分析系統應用之跨領域研發經驗。可進一步瞭解以資料探勘建構知識後，將應用於其他情境互動的過程發展，尋求現場議題問題之現象方法。

本研究承蒙長庚科技大學跨域研究計畫經費補助（計畫編號：ZRRP-F3H0071），在此致謝。

參考文獻

- 丁怡婷、劉志光 (2010)。文字探勘技術應用於中醫診斷腦中風之研究。 *Journal of Data Analysis*, 5 (4), 41-64。
- 翁慈宗 (2009)。資料探勘的發展與挑戰。 *科學發展*, 442, 32-39。
- 張云濤、龔玲 (2007)。資料探勘原理與技術。臺北市：五南。
- 張曉珍 (2013)。運用文字探勘技術在社群行為上之人格預測。未出版之碩士論文，國立交通大學管理學院資訊管理學程，新竹市。
- 梁衍忠 (2012)。資料探勘技術在學生國語學習成就之分析運用。 *教育與多元文化研究*, 7, 65-110。
- 陳世榮 (2015)。社會科學研究中的文字探勘應用：以文意為基礎的文件分類及其問題。 *人文及社會科學集刊*, 27 (4), 683-718。
- 陳昇飛 (2013)。社會互動教學與幼兒數概念之研究。 *朝陽人文社會學刊*, 11 (1), 69-92。
- 童瓊慧 (2008)。資訊檢索結合文字探勘之應用－以中醫婦科專題文獻資料庫為例。未出版之碩士論文，淡江大學資訊與圖書館學系碩士班，新北市。
- 黃仁鵬、張貞瑩 (2014)。運用詞彙權重技術於自動文件摘要之研究。 *資訊管理學報*, 21 (4), 391-415。
- 黃湘萍、趙莉芬 (2018)。應用資訊科技於護理教學的契機與建置規劃。 *源遠護理*, 12 (3), 14-19。
- 簡馨瑩 (2010)。學生提問策略教學對教室裡師生互動與教學序列結構之影響研究。 *當代教育研究季刊*, 18 (3), 125-163。
- Barnes, D. (1976) *From communication to curriculum*. Harmondsworth, England: Penguin
- Ben-Naim, A. (2017). Entropy, Shannon's measure of information and Boltzmann's H-theorem. *Entropy*, 19(2), 48.
- Brigden, D., & Dangerfield, P. (2008). The role of simulation in medical education. *The clinical teacher*, 5(3), 167-170.
- Chao, L.-F., Huang, H.-P., Ni, L.-F., Tsai, C.-L., & Huang, T.-Y. (2017). Construction and application of innovative education technology strategies in nursing. *Hu Li Za Zhi*, 64(6), 26-33.
- Hang, X., Liu, J. N., Ren, Y., & Dai, H. (2005). *An incremental FP-growth web content mining and its application in preference identification*. Paper presented at the International Conference on Knowledge-Based and Intelligent Information and Engineering Systems.

- Johnson, B. K. (2020). Observational experiential learning: theoretical support for observer roles in health care simulation. *Journal of Nursing Education*, 59(1), 7-14.
- Kneebone, R. L., Nestel, D., Vincent, C., & Darzi, A. (2007). Complexity, risk and simulation in learning procedural skills. *Medical education*, 41(8), 808-814.
- Liang, J., Shi, Z., Li, D., & Wierman, M. J. (2006). Information entropy, rough entropy and knowledge granulation in incomplete information systems. *International Journal of general systems*, 35(6), 641-654.
- Sharma, S., Boet, S., Kitto, S., & Reeves, S. (2011). *Interprofessional simulated learning: the need for 'sociological fidelity'*: Taylor & Francis.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*: Morgan Kaufmann.
- Worku, H., & Alemu, M. (2020). Classroom interaction in physics teaching and learning that impede implementation of dialogic teaching: an analysis of student-student interaction. *Bulgarian Journal of Science & Education Policy*, 14(1).



CACET
中華資訊與科技教育學會

A Preliminary Study for Teacher-student Communication Interaction Patterns in Nursing Simulation Education Lesson—Using Text Mining Analysis

Li-Fen Chao (Corresponding author)

PhD, RN, Associate Professor
Clinical Competency Center, Department of Nursing,
Chang Gung University of Science and Technology
Associate Research Fellow
Department of Emergency Medicine, Linkou Branch,
Chang Gung Memorial Hospital, Taoyuan City, Republic of China (Taiwan)
E-mail: lfchao@mail.cgust.edu.tw

Chun-Liang Chen

Adviser
Sport Health Leisure Research Center, SHLRC
National Taipei University of Technology
E-mail: kennyqoo@gmail.com

Hsiang-Ping Huang

PhD, RN, Assistant Professor
School of Nursing of Chang Gung University of Science and Technology
E-mail: hphuang@mail.cgust.edu.tw

Lee-Fen Ni

PhD, RN, Associate Professor
School of Nursing, Chang Gung University of Science and Technology
E-mail: lfni@mail.cgust.edu.tw

Ying-Jung Tseng

PhD, RN, Associate Professor
School of Nursing, Chang Gung University of Science and Technology
E-mail: yjtseng@mail.cgust.edu.tw

Abstract

The classification and definition of classroom teacher-student interaction behavior history has always been limited to the expert's point of view to set qualitative corresponding categories, and after such category settings, there will often be unexplainable or a quantitative indicator that can be applied to teaching. In this research, we hope to use the technology of text mining, analysis and interpretation results by using the research questions formed by multiple research methods of mixed-methods research. Based on the collected information, the information entropy value calculated after the vectorization of the data is used to develop a set of research and teaching aids and an interactive information system for teacher-student interaction record analysis to explore professional core literacy such as care, communication, and logical thinking. Teaching the characteristics of teacher-student interaction and establishing norms. The word segmentation thesaurus that has been established during the implementation process includes 196,392 vocabulary and 99,231 symbols. In the next stage, we will also add academic electronic databases and master's dissertations and other online retrieval archives to conduct text mining processes as a classroom teaching communication interaction. The basis of setting makes the developed system more reliable.

Keywords: mixed-methods research, information entropy, text mining

中華資訊與科技教育學會

