

2026 ICEET

第十五屆 數位學習與教育科技國際研討會
International Conference on E-learning and Educational Technology

智慧教育×永續共榮

論文集 PROCEEDINGS



以拓樸資料分析探索線上學習行為之結構分歧與留存路徑

許舒雅

研究生

東海大學

資訊管理研究所

E-mail : G14490003@thu.edu.tw

余心淳*

教授

東海大學

資訊管理研究所

E-mail : hsyu@thu.edu.tw

摘要

本研究以線上學習情境中學生行為之結構特性為核心，突破既有以退選預測為主之結果導向研究取向，引入拓樸資料分析（TDA）之 Mapper 演算法，探索高維學習行為資料中的分支與連續結構。研究使用 UCI 公開之 OULAD 資料集，選取 CCC-2014J 課程學生平台互動紀錄，建構「週次 × 活動類型」之多維行為向量，並結合 UMAP 嵌入、多尺度覆蓋設計與局部群聚，建立學生行為之拓樸圖模型。結果顯示，學生學習歷程呈現顯著異質性，可穩定區分為早期失聯型、測驗導向無法持續型、穩定留存型與孤島學習型四類結構角色，且退選並非單一路徑，而存在可觀察之過渡節點與行為位移，顯示學習歷程具有動態演變的特性。本研究證實拓樸結構分析可補充傳統預測模型之不足，提供兼具結構詮釋、多尺度穩定性與視覺化優勢之學習行為分析框架，並為後續結合機器學習之可解釋模型奠定基礎。

關鍵字：拓樸資料分析、Mapper 演算法、線上學習行為、學習分析、學生留存與退選

壹、前言

大規模開放式線上課程（MOOCs）的興起及普及，讓學習得以突破學生資格與地域的限制，成為高等教育與終身學習中的重要形式。然而研究發現，線上教學形式經常被批評缺乏真正的社會化互動，因此針對非強烈動機的休閒學習者來說容易失去學習慾望（Baturay, 2015），儘管線上的學習模式在參與人數與課程數量上呈現高度成長，但根據研究指出，線上學習課程長期以來一直存在大量退選問題，且大多數在註冊不久後就選擇停止學習（Reich & Ruipérez-Valiente, 2019），此一現象促使研究者長期關注學生在課程期間的學習行為，期望能及早辨識潛在風險，並設計適當的教學介入措施。

現今相關研究多以「結果導向」為主要分析取向，傾向透過分類模型預測學生最終是否留存，或以平均趨勢描述群體特徵。例如，教育資訊領域的系統性文獻回顧指出，在學生流失議題中，多數研究以邏輯回歸、決策樹與隨機森林等分類模型為核心方法（Silva & Roman, 2021）。然而，此類研究的分析焦點多集中於結果標籤本身，較少從結構層級探討學生在學習歷程中可能經歷之多重行為發展方向與過渡狀態，使得學習歷程的連續性與分歧特徵難以被充分揭示。

為補足上述限制，本文引入拓樸資料分析（Topological Data Analysis, TDA）中的 Mapper 方法，作為重建學習行為空間結構的分析工具。不同於傳統分群方法將學生劃分為互斥類別，Mapper 透過重疊覆蓋與局部連結機制，將高維行為資料轉換為拓樸圖結構，使行為狀態之間的鄰近關係、分支形態與過渡區域得以被視覺化與分析。於此結構中，分支可對應學生學習路徑的不同發展方向，而連接不同分支的節點則可視為行為狀態尚未完全穩定、可能產生策略轉換或投入弱化的過渡區域。本文進一步將此類結構現象連結至學生學習歷程中的行為分化與潛在介入時機，相關概念將於文獻探討中進一步界定，而此結構性使研究者能在不預設行為類型數量或學習階段順序的情況下，探索具有高度異質性與連續變化特性的線上學習行為資料，進而從拓樸形狀理解學習路徑的多樣性及其可能的教學介入意涵。

貳、文獻探討

一、學生留存與流失理論

學習行為分析（Learning Analytics）長期為教育研究關注的焦點，早期理論多以傳統校園情境為背景，強調學生與學術體系及同儕社會環境之整合程度（Tinto, 1975）。然而，隨著教育型態逐漸多元化，特別是成人教育與線上學習環境，學習環境不再以實體校園互動為核心。後來提出的非傳統學生流失模型（Bean & Metzner, 1985）以建立在 Tinto 理論基礎上，針對兼職或通勤的非傳統學生。對非傳統學生來說工作、家庭責任與財務壓力等外部情境因素，往往比傳統意義上的社會整合更能解釋其學習持續性。

延續對非傳統學習者的關注，研究對此進一步提出遠距教育流失的縱貫過程模型，將研究焦點轉向以成人、在職、兼顧家庭與學習的遠距學生，該模型指出，此類型的遠距學生會持續進行「成本效益評估」，於學習壓力與預期成就間權衡，因此課業進度安

排、線上教學品質、評量回饋機制等因素調整學習投入程度 (Kember, 1989)。

綜上所述，近年研究傾向保留 Tinto 所提出「留任為一動態歷程」的理論精神，以回應線上學習情境的實際特性。例如在文獻 (Rovai, 2003) 中指出，雖然 Tinto 的學生留存理論，以及 Bean 與 Metzner 所提出之非傳統學生流失模型，對理解學生留存具有重要貢獻，但兩者皆以實體校園為主要情境，因而不足以解釋線上遠距學生的流失行為。研究認為，學生的留存不僅取決於個人特質與學術表現，更受到外部環境、學習技能與線上教學設計品質的共同影響，因此提供了更貼近成人線上教育情境的解釋框架。

本研究將學生退選視為一種逐步累積的歷程結果，而非單一時間點上的靜態標籤。本文所稱之「結構分歧」，即是指學生在相近學習起點或相似參與狀態下，隨著課程進行逐漸形成不同投入模式、學習策略與退選風險的行為分化現象。而「行為位移」則是指學生行為特徵在學習歷程中由穩定參與、多元互動，逐漸轉向測驗單一導向、表層瀏覽或低互動狀態的變化。此處所稱之位移並非宣稱已找到明確因果性的臨界點，而是將其視為學習投入弱化或策略調整失衡的潛在轉折訊號。透過此一理論脈絡，本研究得以將 Mapper 所呈現之分支、過渡節點與孤島結構，連結至學生學習持續性、參與策略與教學介入時機之討論。

二、教育大數據的可解釋

教育大數據與學習分析的發展，使研究者能透過大量學習行為資料建構模型，以預測學生學習成效與退選風險。現有研究多以分類或預測任務為主要目標，廣泛採用隨機森林、梯度提升樹與深度神經網路等黑盒模型。儘管此類方法在技術層面常能達到較高的預測準確率，卻往往缺乏可解釋性，使其分析結果難以轉化為具體教學介入依據。

在文獻 (Xing & Du, 2019) 中，可解釋性往往比單純的預測效能更為重要；若模型無法說明學生學習困難或退選風險的形成原因，其分析結果便難以轉化為具體的教學決策。類似觀點亦出現在 (Bowers et al., 2013) 早期預警系統的研究中，該研究強調，有效介入策略必須建立在能清楚指出風險來源的分析基礎上，例如低參與度、缺課或作業未完成等具體行為指標，非僅提供抽象的風險分數。

此外，學習行為資料呈現方式亦深刻影響其在教育現場的應用價值。許多研究傾向以整體平均或總量指標描述學生行為，如平均點擊次數、登入頻率或整體互動程度，隱含學生行為具有單一代表型態的假設。(Sedrakyan et al., 2020) 指出，多數學習儀表板僅呈現彙整式績效指標，忽略學習歷程序列性與策略差異，使難以辨識具體的調節需求。

三、拓樸資料分析與結構探索方法

Topological Data Analysis (TDA) 是以拓樸學觀點分析資料形狀的方法論，旨在研究數據的「形狀」，強調在高維與噪音環境下擷取資料全域結構 (Carlsson, 2009)，在探索性資料分析中，Mapper (Singh et al., 2007) 成為 TDA 中具實用性的工具之一，將原本難以觀察的高維點雲簡化為二維或三維的圖形，並保留數據整體的「形狀」，特別適合用於分析具有高度異質性與連續性的行為資料。而 Mapper 與傳統序列分析方法所關注的問題並不完全相同，傳統序列分析方法通常著重於資料隨時間推進的變化規律，

例如事件發生順序、狀態轉移機率、時間點間的相依關係，或特定行為序列對最終結果的影響。此類方法適合回答「學生行為如何隨時間變化」或「哪些時間序列模式可預測退選」等問題。然而，本研究所欲處理的核心問題並非僅是建立時間轉移模型，而是希望從整體學習歷程所形成的高維行為空間中，觀察學生是否呈現不同的結構位置、行為分支與過渡狀態。因此，本文將每位學生的週次與活動類型整合為多維行為向量，再透過 Mapper 檢視學生在整體行為空間中的相對位置與結構關係。

傳統序列分析較重視「時間順序中的變化」，而 Mapper 則較重視「整體行為空間中的形狀與連通結構」。對線上學習行為而言，學生的學習歷程往往不一定呈現單一路徑或線性變化，不同學生可能在相似的初期參與狀態下，逐漸發展出測驗導向、穩定多元參與、表層瀏覽或低互動等不同策略。若僅以平均趨勢或預設類別進行分析，容易忽略這些行為狀態之間的連續過渡與分歧關係。Mapper 透過重疊覆蓋與局部連結機制，使相鄰或部分重疊的行為狀態能被保留下來，因此能呈現學生學習路徑中可能存在的分支、橋接節點與孤島結構。此特性使其可作為傳統預測模型與序列分析之外的互補性方法，用以揭示學生行為異質性與潛在介入位置。

此外，隨著各領域資料量以指數速度成長，資料的高維度與複雜性已成為分析上的主要挑戰。(Madukpe et al., 2025) 對 Mapper 進行跨領域的系統性回顧，指出該方法已在行為軌跡、學習歷程與生物學等領域受到高度關注。在市場行銷領域中，Mapper 被應用在消費者消費行為與複雜市場結構的分析，透過識別數據中的拓撲循環結構，成功揭示了原本看似無關的產品，其實是透過共同購買的品牌作為中介而形成連結，證實了 Mapper 能有效捕捉消費者選擇的連續性與循環性，協助分析人員找到可能存在隱藏的細分群或非線性的「顧客旅程」(Goldfarb, Kwon & Snider, 2016); 在醫療行為領域，Mapper 透過拓撲圖形結構，將深度學習模型內部潛在空間轉換為拓撲圖，使病患分群結構得以視覺化與解讀，提升了深度學習模型的解釋性 (Jurek Loughrey et al., 2024)。

而執行 Mapper 演算法的流程簡述如下 (Singh et al., 2007)：

- (一) 資料預處理：選定資料集進行整理，檢查是否存在缺失欄位。
- (二) 選定濾波函數：將高維資料 X 投影到低維參數空間 $\mathbb{R}$ ，其中 X 為原始資料空間， f 為濾波函數。

$$f: X \rightarrow \mathbb{R} \quad (1)$$

- (三) 參數空間中建立重疊覆蓋：在低維空間建立一組重疊區間或區塊 $\mathcal{U}$ 。

$$\mathcal{U} = \{U_1, U_2, \dots, U_k\} \quad (2)$$

然後拉回到原始資料空間，找出對應的局部子集合。每一個 $f^{-1}(U_i)$ 就是一群在該濾波函數值上相近的資料點。

$$f^{-1}(U_i) = \{x \in X \mid f(x) \in U_i\} \quad (3)$$

- (四) 區塊局部群聚：對每個局部子集合 $f^{-1}(U_i)$ 執行分群演算法，以近似其連通分量，每一個 C_{ij} 是由局部共同行為型態組成。(Singh et al., 2007)。

$$f^{-1}(U_i) \Rightarrow \{C_{i1}, C_{i2}, \dots\} \quad (4)$$

- (五) 建立拓撲圖：每一個 C_{ij} 視為一個節點，若任意兩節點擁有交集則相連。

$$E = \{ (C_{ij}, C_{kl}) \mid C_{ij} \cap C_{kl} \neq \emptyset \} \quad (5)$$

其中 E 代表邊的集合，最終生成的拓模結構即為這些節點集合的神經 (Nerve) 複合體：

$$\text{Mapper}(X, f, \mathcal{U}) = \text{Nerve}(\{C_{ij}\}) \quad (6)$$

參、研究實施與設計

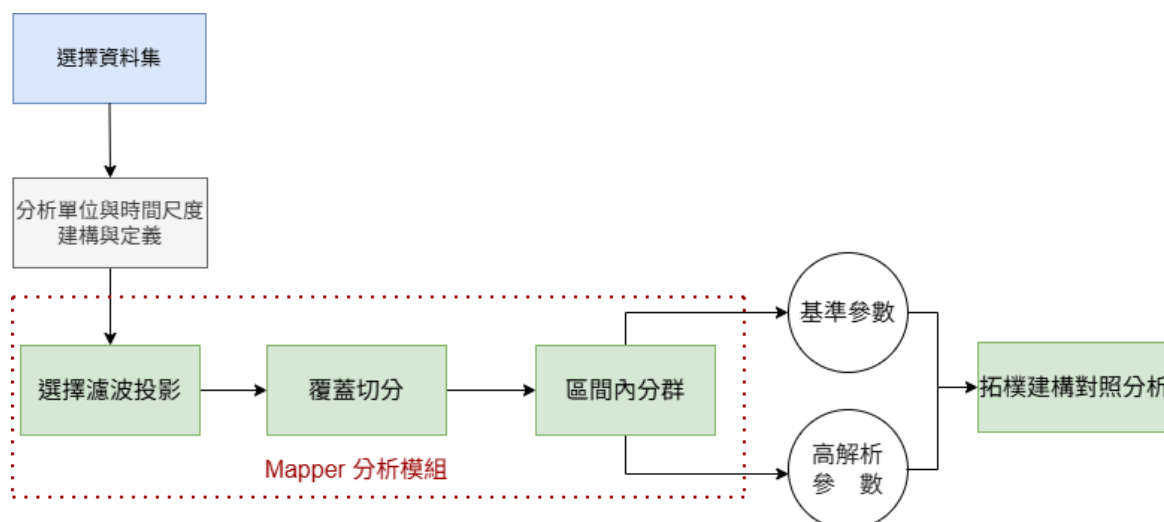


圖 1 研究流程示意圖

圖 1 為本研究的流程架構，詳細方法與設定將於本節後續進一步說明。

一、資料集與研究對象

本文使用 UCI Machine Learning Repository 的 OULAD (Open University Learning Analytics Dataset) 公開教育大數據資料集，在學生學習行為分析領域中具代表性，源自英國公開大學 7 門學科、共 22 個開課期別。資料涵蓋三萬餘名學生的基本背景（性別、年齡）、修課結果（通過或退選）、平台互動紀錄（每日點擊總數）與學習活動類型（如論壇、網頁、測驗）等千萬筆數據。在 OULAD 資料集，每一課程以「課程代號+開課年份期別」為區別，本研究選擇 CCC-2014J 課程作為研究對象，CCC 屬於 STEM（科學、技術、工程、數學）領域的課程，為該課程修課學生人數充足，修課完成及退選的學生有一定的數量，且資料完整性高。研究透過整合學生在平台各區域的點擊次數與活動類別，以此建構每位學生「週次 × 活動類型」的行為向量，以反映學生在學習該堂課的歷程中不同時間點的學習歷程。為了避免僅以學生是否通過或退選的結果標籤作為主導分析，本研究並未預先將學生最終狀態預設特定的行為型態。

二、分析單位與時間尺度建構與定義

本研究將學生行為定義為四類特徵：活動組成比例（包含測驗、內容閱讀及導航比例）用以反映測驗導向、閱讀習慣與導航行為；並結合行為多樣性（活動類型平均值）與結果標記（退選率）評估學習策略廣度及退選風險，時序採用「週 (week)」作為時

間尺度，並以一學期共計 20 週的時間為區間，將學生於課程期間的學習活動依教學週次進行重組。考量原因如下：第一，教育課程課綱設計通常以週為教學單位，小考及課堂作業多以週次進行；第二，設定週單位可以避免時間切割過於零碎，避免資料量過多，在空間分布高度稀疏的問題；第三，相較於以天為單位或是直接比較整體課程行為總和，以週為單位更能清楚反映學生在課程進行中的變化趨勢。研究在建立行為特徵過程中，刻意避免以全體學生的整體平均作為主要分析依據，並未將不同學生的行為進行先試驗整合，不預設學生類型數量，亦不限制學習歷程必須呈現單一路徑，而平均值在本研究僅作為後續節點層級解釋之輔助工具，而非用以代表整體學生學習行為的統一描述。

三、 使用 TDA Mapper 進行計算

本文採用 Mapper 分析流程包含行為向量嵌入、覆蓋設計、局部群聚，以及拓樸圖建構。不同於傳統分群研究，本研究將參數配置視為「觀測尺度」之控制工具，透過多尺度比較以確保結構的穩定性。

(一) 濾波函數與降維：本研究採用 Uniform Manifold Approximation and Projection (UMAP) 作為濾波函數，將高維行為投影至二維空間。因學生行為具強烈非線性特徵，相較於線性的 Principal Component Analysis (PCA)，UMAP 能更有效保留局部拓樸結構，為後續覆蓋切分提供基礎。

(二) 多尺度覆蓋設計：透過設定重疊比例將資料劃分為多個局部區域，確保行為狀態之連續性。本研究將區間數量與重疊率視為解析度控制工具，藉由對比高、低解析度下的拓樸結構，觀察學生行為在不同尺度下的變化特徵。

1. 低解析度模型(為基準參數)：產生結構較為緊密、節點連結度較高的拓樸圖。
2. 高解析度模型：較高解析度設定則有助於放大局部細節，揭露潛在的分支與過渡結構。

(三) 局部群聚與節點建構：本研究採用 K-means 作為局部群聚方法，主要考量其計算穩定、結果易於解釋，且適合作為探索性 Mapper 分析中的基準群聚器。在群聚參數設定上，本研究採用固定值作為各覆蓋區間內的局部切分基準，此設定的目的並非假設每個區間皆必然存在相同數量的學生類型，而是希望在不同局部區域維持一致的分割規則，使後續拓樸圖能反映覆蓋設計與行為空間本身所形成的結構差異。此一機制使 Mapper 不會將學生行為狀態強制切割為完全互斥的類別，而是能保留不同學習策略之間的連續性與過渡關係，後續可進一步以 DBSCAN 或 HDBSCAN 進行敏感度分析，以檢驗主要分支、過渡節點與孤島結構是否在不同群聚方法下仍具穩定性。

四、 解析度調整設計

Mapper 需要設定覆蓋區間數量與重疊比例等參數，因此不同參數組合都會影響拓樸圖的節點數量、分支連結型態清晰度。因此本研究設計解析度調整之對照實驗，觀察學生行為結構在不同解析度下的穩定性與可解釋性。此目的並非尋找「最佳」參數配置，而是使用在常見的解析度範圍內，Mapper 所呈現行為結構是否具有整體型態與

可解釋特徵。

基準參數採用相對較低解析度之覆蓋設計，將嵌入空間切分為數量較少、重疊比例較高之覆蓋區域，建立較大區塊的全域性結構，若此時已展現初步分支，即證實學習路徑具備顯著的結構特徵。

在此基準上，本研究進一步調整為「高解析度設計」，將解析度視為一種觀測尺度，透過增加覆蓋區間數量並調整重疊比例，將行為嵌入空間劃分為更細緻局部分析區域，用以檢查學生行為結構是否會在不同尺度下呈現一致的分化趨勢。若某些結構角色，例如高退選傾向節點、穩定留存分支、過渡節點或孤島結構，在基準解析度與高解析度設定下皆能維持相近的行為特徵，則表示該結構並非單一參數設定下的偶然結果，而具有較高的解釋穩定性。透過此種多尺度比較，本研究可在保留 Mapper 對局部結構敏感性的同時，降低單一覆蓋參數或局部群聚設定對結果詮釋造成的偏誤。

肆、結果與討論

本研究採用 KeplerMapper 實作 Mapper 演算法工具，為開源的 Python 函式庫，提供提供彈性且高效的 Mapper 演算法實作，自動生成互動式 HTML 視覺化介面(van Veen et al., 2019)，圖中節點顏色依據節點內學生之退選比例進行著色，顏色越淺代表該行為狀態中退選風險越高。本文首先以圖 2 左圖基準參數作為結構基準，此設定為解釋在不進行細緻解析度調整情況下，Mapper 對資料所呈現「初始結構樣貌」，初步觀察發現，學生行為空間在低解析度下已展現出明顯的非平均分布，並存在密集的連通分支與結構孤島。然而，由於在此設定下覆蓋區間重疊度較高且數量較少，導致節點間的連結極為稠密，形成了較為混雜的「主幹型」結構。

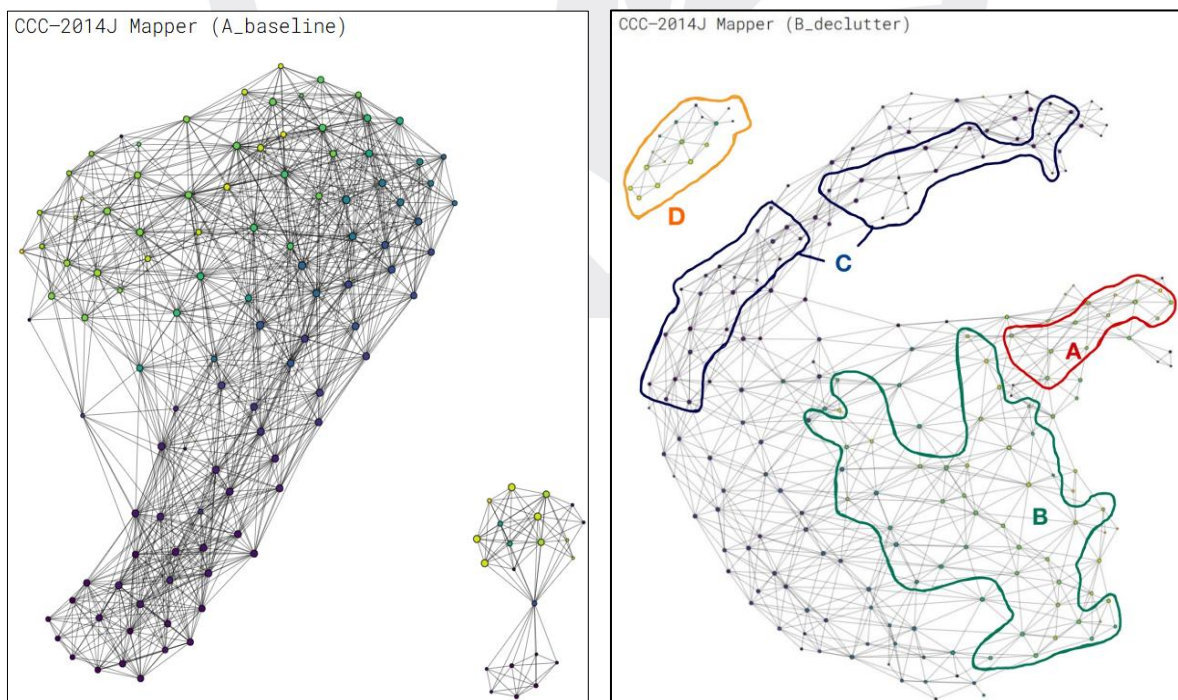


圖 2 Mapper 拓樸結構圖（左為基準參數，右為高解析度參數）

圖 2 右圖為提高解析度並調整覆蓋設計後所產出的拓樸圖，畫面呈現明顯不同的結構型態。相較基礎參數設定，節點數量增加，原先在低解析度下被合併的行為狀態，於較細尺度下自然分化為多個具代表性的局部行為區域。此外，解析度調整後亦可觀察到連結不同分支的過渡節點，反映部分學生行為狀態介於不同策略之間，呈現行為轉換或不穩定參與的特性，顯示解析度調研整對揭露行為歷程連續性的重要性。

本研究將圖 2 右圖結果分為四個學生學習類型，為進一步理解拓樸結構所反映的行為意涵，以表 1 介紹四個學生學習類型具代表性的行為型態，每一類型並未以節點數量作為比較依據，而是各結構角色中符合該學習特徵的拓樸節點來進行行為摘要分析，以避免極端值或單一節點行為對整體結論造成偏誤。

表 1、Mapper 結構角色之節點行為特徵比較

結構角色	退選比例	主要學習活動	早期投入	後期投入	學習互動種類
A. 早期失聯型	高 (>0.80)	表層導覽 停留行為	高 (>0.80)	極低 (<0.10)	極低 (<1)
B. 測驗導向無法持續型	中高 ($0.6-0.9$)	測驗作答 行為	高 ($0.4-0.6$)	低 (<0.20)	中 ($5-7$)
C. 穩定留存型	低 (<0.25)	多元參與 學習活動	中 (>0.30)	高 (>0.30)	高 (≥ 4)
D. 孤島學習型	高度變異 ($0.10-1.0$)	表層導覽 停留行為	近 0	高度變異 ($0-1.0$)	高度變異 ($1-6$)

首先，在 A 類（早期失聯型）中退選比例高於 0.80，為四類結構角色中退選風險最高者，其行為特徵呈現早期集中但在課程中後期即完全消失現象。學生多於課程初期短暫登入平台，但互動行為主要停留於導覽層級（如首頁或子頁面瀏覽），幾乎未進行實質學習活動，整體而言，該類學生的行為多樣性極低，顯示其未能有效進入課程學習脈絡，即快速退出學習。類似的結果為 B 類（測驗導向無法持續型），此類結構角色同樣呈現較高的退選比例，但其行為模式相較於早期失聯型，學生在課程初期具有較高的學習投入，集中於測驗相關活動，顯示以評量為核心的學習策略。然而，該類學生在課程後期的互動明顯下降，未能維持穩定的學習節奏，最終多數學生仍為退選結果。

C 類（穩定留存型）為穩定學習的群體，具有最低的退選比例，此類學生行為分布相對均衡，涵蓋內容閱讀、測驗與其他學習活動，且在課程後期仍維持穩定互動。其行為多樣性顯著高於其他結構角色，顯示學生能以多元方式參與課程並持續投入。而 D 類（孤島學習型）整體節點結構呈現弱及斷裂的狀態，顯示其行為模式與多數學生存在顯著差異且找不到一致規則，且多數學生最終退選，推測這類學生可能存在延遲註冊或本身數據異常，在整體學習行為空間中，此特徵在傳統平均分析中容易被忽略。

除整體分支結構外，本研究亦進一步關注高解析度 Mapper 圖中連接穩定留存型與高退選分支之過渡節點，此類節點的意義不在於其本身已明確屬於留存或退選類型，而在於其節點內學生行為同時具有部分穩定參與與部分風險弱化特徵，呈現介於不同學習策略之間的中介狀態。換言之，過渡節點可被視為學生學習行為尚未完全分化前的結構位置，具有觀察潛在行為位移與教學介入時機之價值。

為避免將 Mapper 圖誤解為單純的時間序列圖，本研究並不將過渡節點解讀為「出現在某一週」的單一時間事件，而是回到節點內學生之週次行為特徵，觀察其風險訊號主要集中於課程哪些階段。具體而言，本研究依據過渡節點內學生之前期、中期與後期行為摘要，整理其可能風險訊號，包括內容閱讀比例、測驗活動比例、導航行為比例、活動多樣性與總互動量等指標。若過渡節點中的學生在課程前期仍保有一定互動，但於中期開始出現活動類型單一化、內容閱讀下降或後期互動快速減少，則可將該階段視為潛在的預警時間窗。此分析使過渡節點不僅是拓撲圖中的橋接位置，也能進一步轉化為教學現場可理解的風險訊號。

表 2、過渡節點之各階段行為特徵與潛在預警意涵

可觀察行為階段	過渡節點可能風險訊號	教學介入意涵
課程前期	仍有登入、瀏覽行為，但活動多樣性不足。	學生尚未完全失聯，但可能尚未建立穩定學習策略。
課程中期	測驗活動仍存在，但內容閱讀與其他互動下降。	學習可能轉為短期應付評量，需提供策略提醒。
課程中後期	總互動量下降，後期投入不足。	可能逐漸流向高退選分支，可作為預警時間窗。
常見於學習轉弱階段	導航或頁面瀏覽比例較高，實質學習活動不足。	可透過課程導引、學習提醒或資源推薦進行介入。

如表 2 所示，從教學介入角度而言，過渡節點所呈現的價值在於：相較於早期失聯型學生已接近完全退出學習，過渡節點中的學生仍保留部分平台互動，因此可能仍有較高的介入可能性。例如，若學生仍持續參與測驗但內容閱讀與活動多樣性逐漸下降，教師可針對其學習策略提供提醒或補充資源；若學生由多元參與轉向表層導覽，則可透過學習進度提醒、課程導引或同儕互動設計，降低其進一步流向高退選分支的可能。由此可見，過渡節點的教育意涵並不僅在於描述學生類型，而在於協助辨識學習投入由穩定轉向弱化的過程，進一步提供比最終退選預測更早期且更具行動性的介入依據。

伍、 結論與未來展望

本文以線上學習情境中學生行為探索性分析為核心，透過 Mapper 方法揭露線上

學習情境中學生行為結構的實際分歧型態，證實學生在線上課程中的學習行為並非可由單一平均模型所代表，而是由多條具結構差異的行為路徑所構成，行為異質性本身即為理解退選與留存的重要分析單位，顯示即使學生的最終結果為退選，在高解析度拓撲圖中，仍可觀察到連接穩定留存型與高退選分支的過渡節點，顯示學生可能在課程進行中產生可被追蹤的行為位移，提供教學介入的時間點與訊號依據。本研究並非為取代既有之退選預測研究為目標，相反地，本研究提出的是一種互補性的分析視角，強調從結構層級理解學習行為的多樣性與可能路徑，以補充現有以效能或平均趨勢為導向的研究取向。基於此定位，本研究主要聚焦於學習平台的點擊與活動參與行為，未來可納入成績表現、作業回饋或討論文本內容等更多元的資料來源。透過結合量化行為特徵與語意或表現相關資訊，Mapper 所建構的拓撲結構有機會反映更細緻的學習歷程差異，並提升節點層級行為詮釋的深度。此外，後續研究亦可探索將結構性分析結果結合機器學習，將 Mapper 所揭露之拓撲結構作為一種模型外部的結構性解釋層，兼具預測效能與理論詮釋能力之研究架構，強化學習分析模型在教學介入與學習支持決策上的實用價值。

參考文獻

- Baturay, M. H. (2015). An overview of the world of MOOCs. *Procedia - Social and Behavioral Sciences*, 174, 427–433.
- Bean, J. P., & Metzner, B. S. (1985). A conceptual model of nontraditional undergraduate student attrition. *Review of Educational Research*, 55(4), 485–540.
- Bowers, A. J., Sprott, R., & Taff, S. A. (2013). Do we know who will drop out?: A review of the predictors of dropping out of high school: Precision, sensitivity, and specificity. *The High School Journal*, 96(2), 77–100.
- Carlsson, G. (2009). Topology and data. *Bulletin of the American Mathematical Society*, 46(2), 255–308.
- da Silva, J. J., & Roman, N. T. (2021). Predicting dropout in higher education: A systematic review. In *Proceedings of the Simpósio Brasileiro de Informática na Educação (SBIE)* (pp. 1107–1117).
- Goldfarb, A., Kwon, J. B., & Snider, T. (2016). *Detecting potential product segments using topological data analysis* (Working paper).
- Jurek-Loughrey, A., Gault, R., Ahmaderaghi, B., Fahim, M., & Bai, L. (2023, November). Interpreting deep patient stratification models with topological data analysis. In *International Conference on e-Health and Bioengineering* (pp. 563–574). Cham, Switzerland: Springer Nature Switzerland.
- Kember, D. (1989). A longitudinal-process model of drop-out from distance education. *The Journal of Higher Education*, 60(3), 278–301.
- Madukpe, V. N., Ugoala, B. C., & Zulkepli, N. F. S. (2026). A comprehensive review of the mapper algorithm, a topological data analysis technique, and its applications across various fields (2007–2025). *International Journal of Data Science and Analytics*, 21(1), 56.

- Reich, J., & Ruipérez-Valiente, J. A. (2019). The MOOC pivot. *Science*, 363(6423), 130–131.
- Rovai, A. P. (2003). In search of higher persistence rates in distance education online programs. *The Internet and Higher Education*, 6(1), 1–16.
- Sedrakyan, G., Malmberg, J., Verbert, K., Järvelä, S., & Kirschner, P. A. (2020). Linking learning behavior analytics and learning science concepts: Designing a learning analytics dashboard for feedback to support learning regulation. *Computers in Human Behavior*, 107, 105512.
- Singh, G., Mémoi, F., & Carlsson, G. E. (2007). Topological methods for the analysis of high dimensional data sets and 3D object recognition. In *Proceedings of PBG@ Eurographics* (Vol. 2, pp. 91–100).
- Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, 45(1), 89–125.
- Van Veen, H. J., Saul, N., Eargle, D., & Mangham, S. W. (2019). Kepler Mapper: A flexible Python implementation of the Mapper algorithm. *Journal of Open Source Software*, 4(42), 1315.
- Xing, W., & Du, D. (2019). Dropout prediction in MOOCs: Using deep learning for personalized intervention. *Journal of Educational Computing Research*, 57(3), 547–570.

Exploring Structural Divergence and Retention Pathways in Online Learning Behaviors Using Topological Data Analysis

SHU YA XU

Student

Graduate School of Information Management

Tunghai University

Taichung City, Taiwan

E-mail : G14490003@thu.edu.tw

HSIN CHUN YU*

Professor

Graduate School of Information Management

Tunghai University

Taichung City, Taiwan

E-mail : hsyu@thu.edu.tw

Abstract

This study investigates the structural characteristics of student learning behaviors in online learning environments within the field of learning analytics, moving beyond outcome-oriented dropout prediction. Mapper-based topological data analysis is used to describe branching and continuity in high-dimensional activity traces. Using the Open University Learning Analytics Dataset (CCC-2014J), each learner is encoded as a “week $\times$ activity type” interaction vector. UMAP serves as the lens; a multi-scale cover with local clustering yields a behavioral graph, and role stability is examined across cover resolutions. The map indicates pronounced heterogeneity and four recurrent roles: early disengagement, assessment-driven but short-lived participation, sustained engagement, and isolated activity. Dropout follows multiple transitions, with intermediate nodes marking behavioral shifts over time. This structural view complements conventional models and supports the development of structure-aware features for explainable learning analytics.

Keywords: Topological Data Analysis, Mapper Algorithm, Online Learning Behaviors, Learning Analytics, Student Retention and Dropout

