

基於匿名評論的課程推薦系統

學生：曾國維

指導教授：劉立頌 教授

摘要

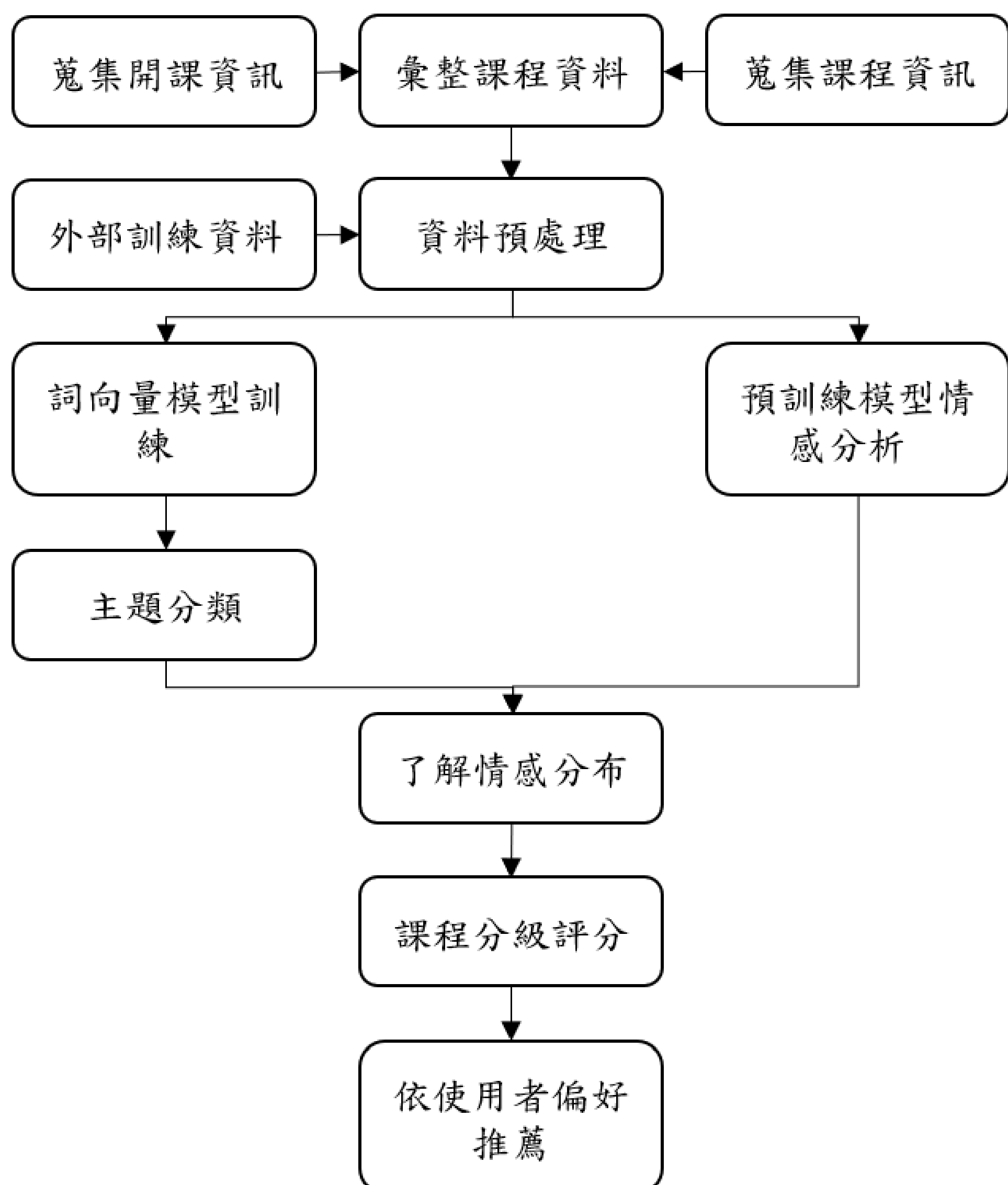
如何在資訊爆炸的現代建立有效推薦系統了解使用者並推薦適合的產品日益重要。然而多數研究需要網站中評分和評論者身份才得以分析，因此本研究想探討僅靠匿名評論建立有效推薦系統的可能性。

實驗流程與架構

本研究使用到的資料是透過網路爬蟲抓取使用傳統HTML技術的國立中正大學排選課系統中的網頁原始碼，以及採用XML(AJAX)的課程評論網站CCU PLUS中所接收的JSON資料。最後將評論做翻譯和去除雜訊後即可進行後續的分析計算。

將評論劃分為“評分”、“難度”、“負擔”三大主題，並以“課程”為額外全面參考主題。利用Tomas Mikolov團隊提出的Skip-gram詞嵌入模型行透過非監督式學習訓練詞向量，再使用餘弦相似度分析每則評論所歸屬的主題。

將評論輸入預訓練中文情感模型後輸出一顆星到五顆星的機率(越多星越正面)，但每個人用詞習慣不同，機率代表的正負面程度難以具有參考價值，於是進一步歸納成正面、負面和中立。



圖一、實驗流程圖

實驗結果與分析

了解每堂課程在不同主題下“正面”和“負面”情緒之評論數量並記錄。下圖為紀錄分析結果的CSV檔，“主題_p”表示該主題正面評論數量，“主題_n”表示該主題負面評論數量。

| name | id | 評分_p | 難度_p | 負擔_p | 課程_p | 評分_n | 難度_n | 負擔_n | 課程_n |
|--|---------|------|------|------|------|------|------|------|------|
| 文學概論(二) Introduction to Literature (II) | 1101812 | 1 | 9 | 2 | 1 | 0 | 1 | 1 | 0 |
| 語言哲學Philosophy of Language | 1252306 | 2 | 3 | 2 | 4 | 1 | 0 | 0 | 1 |
| 線性代數Linear Algebra | 5201007 | 4 | 6 | 3 | 3 | 0 | 0 | 1 | 1 |
| 初級會計學(二) Accounting (II) | 5261105 | 2 | 2 | 4 | 1 | 0 | 1 | 1 | 0 |
| 數學的過去與未來The Past and Future of Mathematics | 7201002 | 4 | 7 | 1 | 7 | 0 | 2 | 0 | 1 |

圖二、實驗分析結果(部分)

統計不同主題下正負面情緒評論的分布，根據每個群體的狀況將課程在不同主題情緒做分級，最後把各主題的正面分級扣除負面分級後乘上偏好權重得出對課程的評分，由高至低列出即為推薦結果。

| 主題 | 正面評論分級 | 負面評論分級 |
|----|---|---|
| 評分 | Level 1 : num<=1 Level 2 : 1<num<=3 Level 3 : num>3 | |
| 難度 | Level 1 : num<=2 Level 2 : 2<num<=5 Level 3 : num>5 | Level 1 : num=1 Level 2 : num=2 Level 3 : num<2 |
| 負擔 | Level 1 : num<=1 Level 2 : 1<num<=2 Level 3 : num>2 | |
| 課程 | Level 1 : num<=1 Level 2 : 2<num<=4 Level 3 : num>4 | |

表一、評論分級

為了驗證系統輸出是否符合要求，下面以主題“難度”做為測試範例，將“難度”的權重設為1，其他主題的權重設為0，並觀察輸出結果的最高分項是否明顯包含該主題的正面評論。下圖為被推薦課程在評論網站中的部分言論，可以發現本系統能確實依照設定得出所求。

老師教得還不錯，如果不想聽課也能自己讀
常點名，會傳簽到單
作業不難，能輕鬆拿100
考試偏簡單，只要看過幾遍投影片就穩了

老師很熱情，課程內容也不會很難
很多東西之前的課就有碰過，然後更加延伸居多

圖三、受推薦課程留言

結論

現階段研究已經能滿足一定程度的推薦需求，然而過程中也發現到許多能增進使用者體驗的設計方向，例如：分析課程大綱以便推薦相似或多元的課程、了解不同授課教授之間的相似或相異程度作為推薦參考等。期許日後的研究能更加完善。