認知神經科學研究所
Institute of Cognitive Neuroscience
National Central University, Taiwan
國立中央大學生醫科學與工程學系
Department of Biomedical Sciences and Engineering

教授講得開心 學生聽得舒心

以協同過濾分析設計之課程推薦系統

ACL2020 陳品瑋 羅珮瑄 陳華婕
action & cognition laboratory

研究動機

學校課程眾多，學生無法單依課程名稱選出有把握獲得高分的課程。因此期許能透過課程推薦系統，以系統預測之課程分數作為學生的選課參考。

研究目的

- 由100~106學年的學生成績資料庫，比較以模型為基礎 (model-based) 以及以記憶為基礎 (memory-based) 的協同過濾 (collaborative filtering) 在成績預測上的準確度。
  - Model-base: 潛在特徵模型 (latent factor model) 中的 BRISMF (Biased Regularized Incremental Simultaneous Matrix Factorization)。
  - Memory-base: 各種基線得推薦算則 (Baseline recommender) 。

研究方法

基線 (Baseline)

基線 (Baseline)	計算公式	預測分數
Global Average	$\widehat{r_{ui}} = \mu$	全部資料的平均值
User Average	$\widehat{r_{ui}} = \mu_u$	使用者u資料的平均值
Item Average	$\widehat{r_{ui}} = \mu_i$	項目i資料的平均值
User-Item Baseline	$\widehat{r_{ui}} = \mu + b_u + b_i$	由Global Average加上 使用者u與項i目的傾向

Latent factor model

假設所有潛在特徵 (latent factors) 各自代表一個維度，組成一個潛在特徵空間 R。每個使用者 u 與項目 i 在 R 中都是一個向量，以使用者向量 ρ_u 與項目向量 q_i 兩者之內積（表示使用者與項目在空間上的相似性）預測使用者對項目的評分 $\widehat{r_{ui}}$：

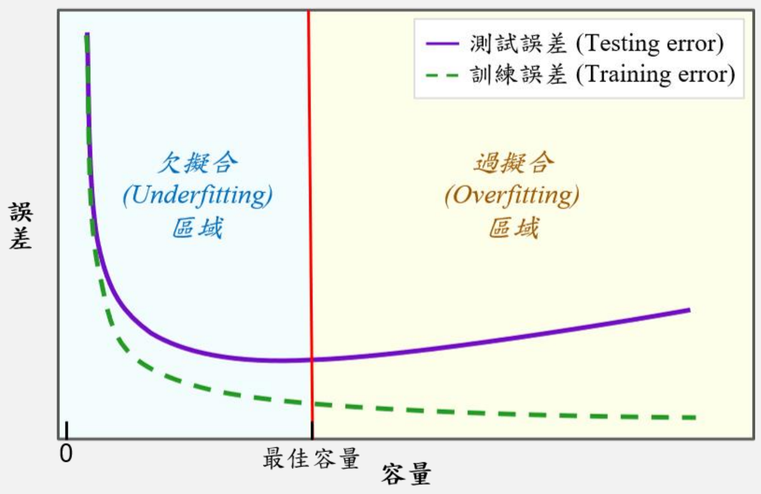
$$\widehat{r_{ui}} = q_i^T \rho_u$$

BRISMF

最佳化機器學習 (Machine learning) 模型

欠擬合 (Underfitting)
模型的訓練誤差無法降低。

過擬合 (Overfitting)
雖然模型的訓練誤差很低，但對未觀測資料的預測表現很差。



*容量 (Capacity): 模型擬合各種函數的能力。

在損失函數中 (Loss function) 加入懲罰項以避免模型過擬合:

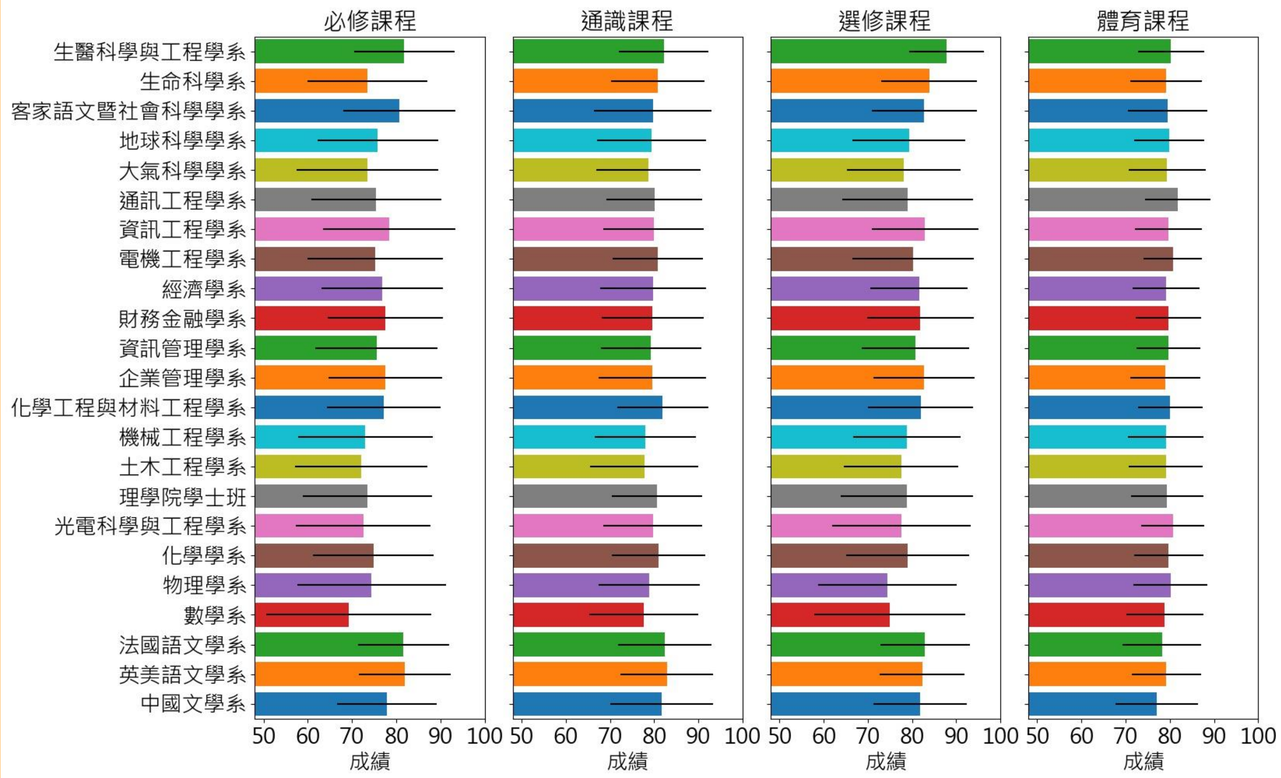
$$\min_{p*,q*,b*} \sum_{(u,i\in k)} (r_{ui} - \mu - b_u - b_i - q_i{}^T p_u)^2 + \lambda(b_u{}^2 + b_i{}^2 + \|p_u\|^2 + \|q_i\|^2)$$

資料篩選

- 大學部學生
- 課程屬性：必修、通識、選修、體育
- 成績：排除空值與零分
- 排除資料數量少於五筆的學生與課程

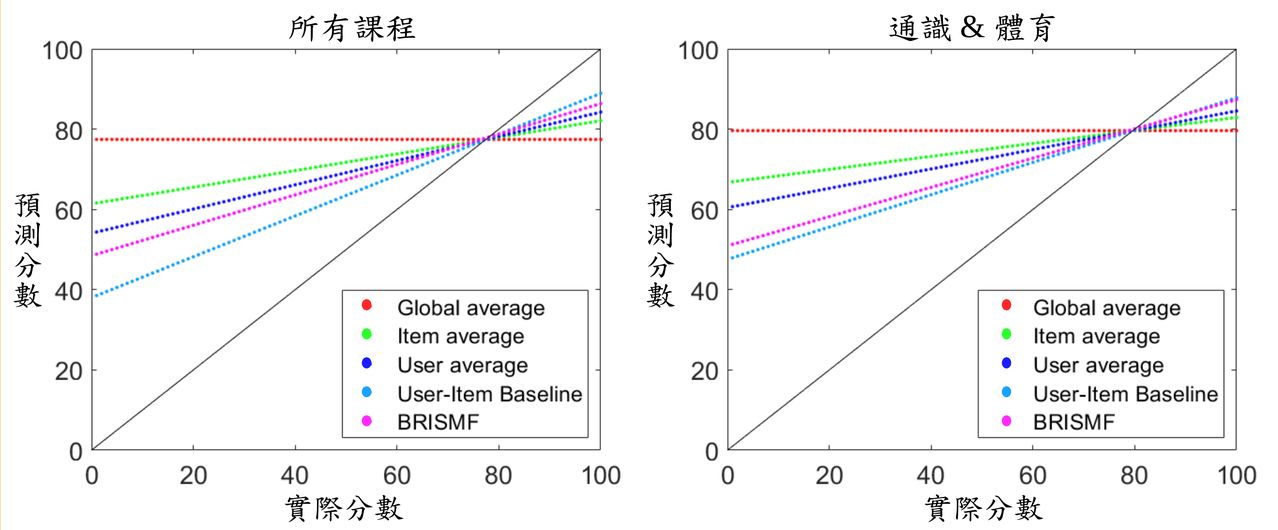
資料數量	原始資料	篩選後資料
總資料	867400筆	486139筆
學生	36344名	14253名
課程	4275堂	1756堂

各系學生課程平均成績



研究結果

以所有學生資料訓練之BRISMF模型與基線比較預測結果

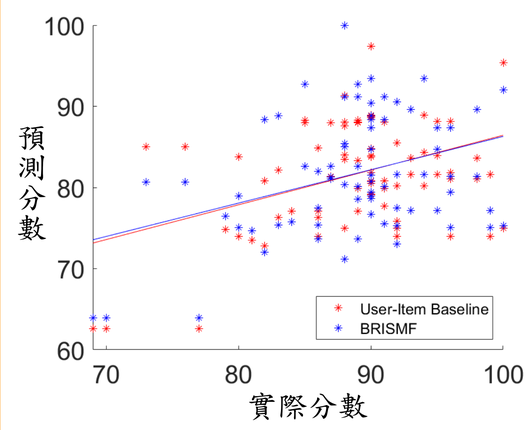


均方根誤差 (Root-mean-square error, RMSE)

訓練與預測資料範圍	Global average	Item average	User average	User-Item Baseline	BRISMF
所有課程	13.48	12.01	11.27	10.16	10.48
通識、體育	10.07	9.22	8.77	7.99	8.21

測試 (Testing) 非資料庫資料

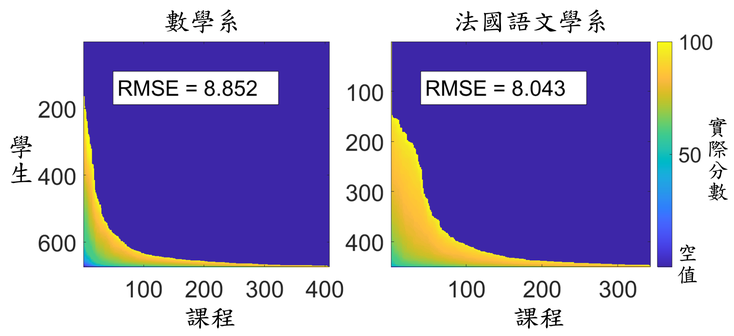
RMSE
User-Item Baseline: 10.31
BRISMF: 10.73



各系 (共 23 個) 之RMSE 與實際分數資料之空洞程度的相關分析

實際資料	Pearson 相關係數	*p (雙尾)*
總數	0.277	0.202
學生-課程矩陣之空洞程度	0.368	0.084

空洞程度最大 & 最小的實際分數資料



討論與結論

- 目前的 BRISMF 模型之訓練資料數不夠多且結構鬆散 (學生-課程矩陣空洞程度之平均值為88.45 %)，預測效果不如以學生平均分數及課程平均分數做出的成績預測。
- **未來展望:** (1) 增加模型複雜度，考慮更多因素 (如: 學期、系級等)，降低預測誤差。(2) 結合人性化使用介面，提供網路系統供學生作為選課參考使用。

參考資料

1. 曹秀丽. (2018). 协同过滤技术在高校选课推荐系统中的应用. *吉首大学学报(自然科学版)*, *39*(1), 34-39.
2. Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. Computer, 42(8), 30-37.
3. Orange3-Recommendation Documentation Release 1.0.0 Salva Carrión

- **指導老師:**
  張智宏 Erik Chihhung Chang
  ✉ audachang@gmail.com