

1. 離線評估

張子恩

September 10, 2025

1 評估目的與重點 (Baseline)

本章建立一套基準評估流程 (baseline)，以「已知委員」之歷史審查資料作為參考集，不以重現人工名單之準確率為目標，而用於檢驗候選人推薦演算法的合理性、穩定性與實用性，並作為後續 (如 seek 版) 與他法比較的共同基準，亦可據此提供實務建議。

- 合理性 (是否把參考人選推到前段)：採用 HitPosition 與 Success@K (命中位置 $\leq K$ 記 1，否則 0)，檢查參考委員是否落在合乎題意的前段名次。
- 穩定性 (不確定性大小)：以 bootstrap 取得 95% CI、變異數、CV (標準差/平均) 與相對 CI 寬度 (CI 長度/平均)。數值越小代表估計越穩定，可用於跨學門/不同 K 比較。
- 實用性 (不同寬鬆度下的覆蓋)：採 K-grid ($K = 10, 20, \dots, 100$) 觀察覆蓋曲線，作為 UI 顯示範圍與流程門檻設定依據。
- 對照與建議：在共同指標下與其他演算法 (如 seek) 比較 (曲線、AUC of Success@K、前段命中占比)，據以提出資料/特徵/重排與門檻之調整建議。

本研究使用以下統計指標以評估推薦系統的排序表現與穩定性：

- **K 的選取**：雖然以母體粗估 (約 50,000 名專任師資與研究人員、85 學類) 推得每學類規模約 $n \approx 588$ ，但實證顯示 $K = 100$ 時 Success@K 約 70%，且在 $K \in [80, 100]$ 已明顯趨緩，顯示再擴大名單只帶來有限提升，卻增加人工審視成本。
- **Success@K**：表示在推薦清單長度為 K 時，是否成功命中目標已知委員。若第一位已知委員出現在前 K 名推薦名單中，則視為命中。可視為命中率隨推薦長度的累積表現。
- **命中位置 (Hit Position)**：成功命中時，目標委員在推薦清單中的實際排名位置。平均命中位置反映整體排序效果，中位數則避免極端值干擾。
- **Bootstrap 指標**：透過重抽樣 (bootstrap) 方式估計不確定性，除了常見的 95% 信賴區間外，亦延伸計算以下相對化統計量：

– 95% 信賴區間 (CI)：由重抽樣分布計算所得的範圍，區間越窄代表估計越穩定。

– 變異係數 (CV)：

$$CV = \frac{sd_{boot}}{mean} \times 100\%,$$

衡量平均值的相對變異程度；數值愈小代表估計愈穩定。

– 相對 CI 寬度 (Relative CI Width)：

$$RelWidth = \frac{CI_{hi} - CI_{lo}}{mean} \times 100\%,$$

以 CI 長度相對於平均值比例來量化不確定性，方便不同學門或不同 K 間比較。

– 解讀建議：一般而言，

* $CV < 10 \sim 15\%$ ，或相對 CI 寬度 $< 20\%$ ，代表估計結果相對穩定；

* $CV > 30\%$ ，或相對 CI 寬度 $> 40\%$ ，則顯示不確定性偏高，需謹慎解讀，可能與樣本過少或命中率過低有關。

- **命中比例 (Hit Rate)**：指成功命中的題數佔總題數之比例。亦即 Success@100 的結果，可作為最終命中上限表現的指標。

2 資料處理

本研究使用之已知資料包含 topA_109 至 topA_114 檔案，前五筆如圖 1，首先進行合併後，需將題目原本國科會分類學門對應至碩博論文學門分類。由於原始資料中的「申請學門」屬於國科會內部代碼，需透過人工建立對應表 subject_map，將其轉換為統一的學門名稱。對應規則如下：

△ 申請人	△ 申請學門	△ 中文題目	△ 英文題目	△ 中文摘要	△ 英文摘要	△ 審查委員姓名	△ 審查委員email	△ 派案情形
遺漏: 0% 相異: 8%	遺漏: 0 (0%) 相異: 19 (2%)	遺漏: 0% 相異: 8%	遺漏: 0 (0%) 相異: 90 (8%)	遺漏: 0 (0%) 相異: 90 (8%)	遺漏: 0 (0%) 相異: 90 (8%)	遺漏: 0 (0%) 相異: 152 (14%)	遺漏: 0 (0%) 相異: 152 (14%)	遺漏: 0 (0%) 相異: 152 (14%)
90 相異值	H14 政治學 H06 歷史學 H23 藝術學 其他 60%	90 相異值	90 相異值	90 相異值	90 相異值	152 相異值	152 相異值	同意
0 洪志清	H40 財金及會計	行為財務金融的	3 Essays on Behav	根據金融理論，每	While fundamenta	葛博	myger@narlabs.org.tw	同意
1 洪志清	H40 財金及會計	行為財務金融的	3 Essays on Behav	根據金融理論，每	While fundamenta	吳宏德	dshand5+W01@gmail.com	同意
2 鄭子含	H12 心理學	人類對音樂及語	Behavioral and Ne	時間知覺是人類與	Time perception is	曹峰銘	tsaosph@ntu.edu.tw	同意
3 洪肇基	H11、H12 心理學	學齡注意力不足	Cross Cultural Insi	先前文獻分別就日	Background and a	曹峰銘	tsaosph@ntu.edu.tw	同意
4 陳嘉峰	H12 心理學	正向親職參與及	Positive parenting	此論文旨在揭露於	Objectives: The pr	曹峰銘	tsaosph@ntu.edu.tw	同意

Figure 1: 原始資料前五筆

Table 1: 國科會學門與碩博論文學門分類對應表

國科會學門	碩博論文學門
H23 藝術學	藝術學類
H06 歷史學	歷史學類
H14 政治學	政治學類
H13 法律學	法律學類
H22 區域研究及地理	地理學類
H17 社會學	社會
H12 心理學	心理學類
H11 教育學	教育
HSS01 數學教育	教育
HSS03 資訊教育	教育
H09 人類學及族群研究 (含客家研究、原住民研究)	人類學學類
H01 文學 (中國文學、台灣文學、客家文學、原住民文學等)	文學
H05 文學 (二) (外國文學)	外國語文學類
H15 經濟學	經濟學類
H19 傳播學	傳播學類
H41 管理學一	管理
H42 管理學二	管理
H40 財金及會計	財務金融學類
HA2 體育學	體育學類
HA3 圖書資訊學	圖書
HV1 新興/其他領域	未分類
HSS07 科技、社會與傳播	未分類
HSS05 醫學教育	未分類

上述轉換確保所有題目能以國科會學門分類進行分析，其中未能對應的題目則保留為「未分類」，以利後續搜尋與統計。

3 評估結果分析與可視化

為了全面評估本系統在推薦前三位審查委員時的命中表現與推薦品質，將從下列三個面向進行分析與視覺化呈現：

1. **Success@K** 曲線：透過繪製成功命中曲線，檢視不同推薦列表長度 K 下的命中比例變化，藉此反映系統推薦在前 K 位中的整體表現趨勢。

2. 命中位置分布：進一步分析每位題目對應的第一位命中委員、第二位命中委員等實際出現位置，並透過統計指標（如平均命中位置、中位數、Bootstrap 信賴區間等）量化系統精準度。
3. 依學門分析：將題目依照所屬學門進行分組，探討推薦系統在不同學門下的推薦效果是否存在顯著差異，作為未來調整推薦模型權重或學門特定處理的參考依據。

本評估的分析僅聚焦於前三位推薦委員。原因在於：雖然系統可產出完整的 Top-100 推薦名單，但根據標註資料，已知委員在第四名以後的命中情形極為稀少，且位於前 100 名以內的題目數量亦不足，無法提供統計上穩健的估計（可參考表 3）。基於此考量，本評估不進一步討論前三名以後的推薦結果。

3.1 第一位命中 Success@K 單獨觀察

圖 2 顯示針對第一位已知委員的 Success@K 曲線。可見當推薦清單長度達 $K = 50$ 時，命中率已達 57.14%，進一步擴增至 $K = 80$ 時則有 67%，第一位關鍵委員皆有在推薦系統前 100 名內。表 2 彙整不同 K 值下的 Success@K 數據。

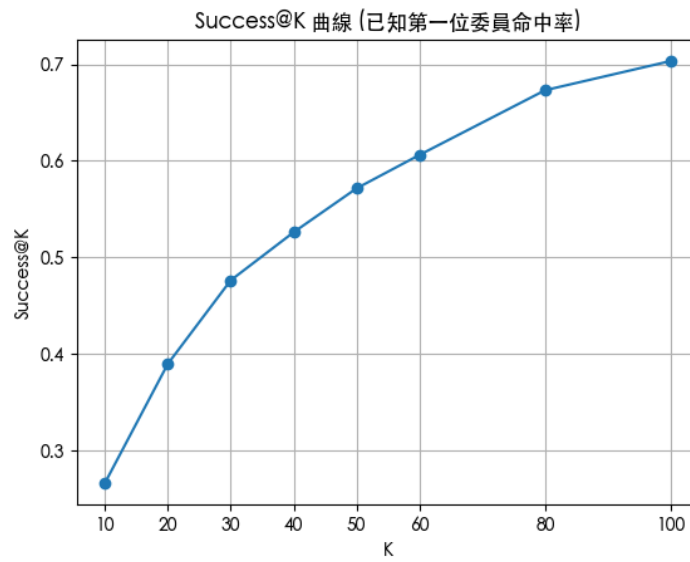


Figure 2: 第一位已知委員命中 Success@K 曲線

Table 2: 不同 K 值下的 Success@K 命中率（第一位已知委員）

K 值	Success@K (命中率)
10	26.62%
20	38.96%
30	47.62%
40	52.60%
50	57.14%
60	60.61%
80	67.32%
100	70.35%

3.2 前三位命中 Success@K 曲線與統計概覽

表 3 彙整系統在全部 462 題中，對前三位已知委員的命中統計結果。

Table 3: 前三位已知委員命中數與比例統計			
命中位次	未命中題數	總題數	命中比例
第 1 位委員	137 題	462 題	70.35%
第 2 位委員	250 題	462 題	45.89%
第 3 位委員	326 題	462 題	29.44%

由表中可知：

- 第一位已知委員的命中比例最高，達 70.35%；
- 第二位與第三位委員的命中比例分別為 45.89% 與 29.44%，呈現遞減趨勢；

需特別說明的是，人工指派本身就可能受多項考量與主觀偏好影響，而推薦系統則是依演算法邏輯給出排序推薦，兩者之間難免存在差異，屬合理現象，因此不宜將命中率高直接對應至「重要性排序」之準確與否。圖 3 則呈現不同 K 值下，推薦清單中成功命中前三位已知委員的 Success@K 曲線。其中：

- 第 1 位委員（藍線）在 $K = 100$ 時命中率可達 70.3%；
- 第 2 位委員（橘線）與第 3 位委員（綠線）則分別為 45.9% 與 29.4%；

此圖清楚呈現系統在清單長度擴增後的命中效果，且始終維持第 1 位委員的預測最為準確，具實務應用價值。

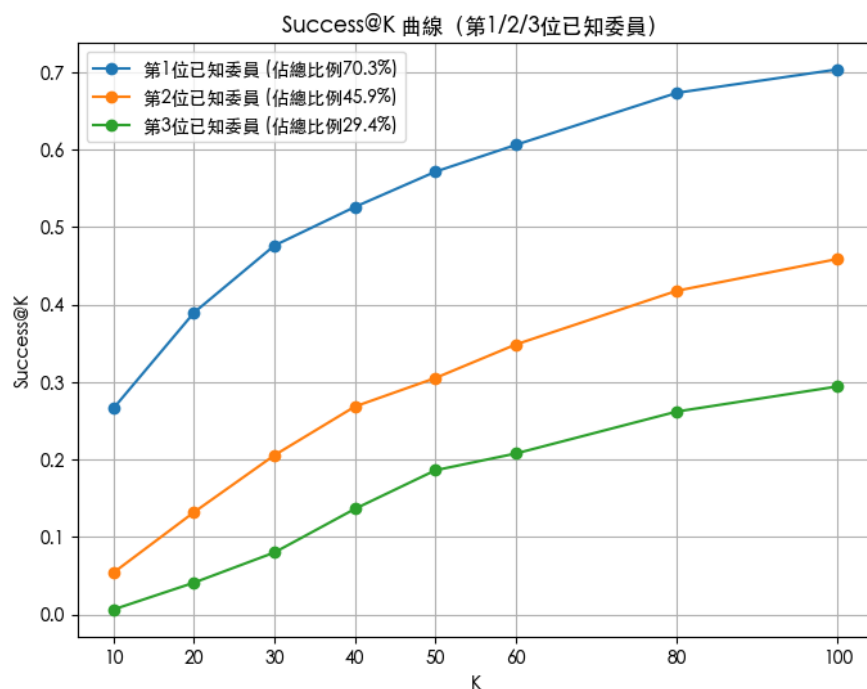


Figure 3: 前三位已知委員 Success@K 曲線

3.3 前三位委員 Success@K 與 Bootstrap 95% 信賴區間

為驗證 Success@K 命中率的穩定性與模擬真實情況，我們以 $n = 462$ 題為基礎進行 bootstrap 重抽樣 ($B = 5000$ 次)，針對第 1 至第 3 位已知委員，計算在不同推薦清單長度 K 下的 Success@K 平均值及其 95% 信賴區間。結果如圖 4 及表 4 至表 6 所示。重點如下：

- 覆蓋率：於 $K = 100$ 時，三者 Success@K 約為第 1 位 0.703、第 2 位 0.459、第 3 位 0.294；在 $K = 50$ 時分別約為 0.571、0.305、0.186，呈遞減趨勢。
- 不確定性：第 1 位在各 K 的區間較窄、CV 較小（穩定）；第 2 位次之；第 3 位在小 K 時 CI 與 CV 較大（樣本稀少所致），但隨 K 增加而收斂（見表 4 表 6 之 CI 寬度、CV 與相對寬度）。
- 由於第一位第二位已知委員的 Success@K 曲線在不同 K 下皆呈現相對穩定的命中表現，因此在比較不同演算法時，可以比較第一位第二位委員的命中情形作為主要指標。

Table 4: 第 1 位委員 Success@K 統計摘要 (Bootstrap 95% CI, $n = 462$, 未命中率 29.65%)

K	Mean	95% CI	CI 寬度	Std	CV%	相對寬度%
10	0.266	[0.227, 0.307]	0.080	0.020	7.67%	30.08%
20	0.390	[0.346, 0.435]	0.089	0.023	5.81%	22.79%
30	0.476	[0.431, 0.524]	0.093	0.024	4.99%	19.55%
40	0.526	[0.483, 0.571]	0.089	0.023	4.30%	16.87%
50	0.571	[0.526, 0.617]	0.091	0.023	4.06%	15.91%
60	0.606	[0.563, 0.649]	0.087	0.022	3.64%	14.29%
80	0.673	[0.630, 0.714]	0.084	0.022	3.20%	12.54%
100	0.703	[0.660, 0.745]	0.084	0.022	3.06%	12.00%

Table 5: 第 2 位委員 Success@K 統計摘要 (Bootstrap 95% CI, $n = 462$, 未命中率 54.11%)

K	Mean	95% CI	CI 寬度	Std	CV%	相對寬度%
10	0.054	[0.035, 0.076]	0.041	0.010	19.39%	76.00%
20	0.132	[0.102, 0.165]	0.063	0.016	12.13%	47.54%
30	0.206	[0.169, 0.242]	0.074	0.019	9.13%	35.79%
40	0.268	[0.227, 0.310]	0.082	0.021	7.82%	30.65%
50	0.305	[0.264, 0.346]	0.082	0.021	6.88%	26.95%
60	0.348	[0.305, 0.392]	0.087	0.022	6.34%	24.84%
80	0.418	[0.372, 0.461]	0.089	0.023	5.42%	21.24%
100	0.459	[0.413, 0.504]	0.091	0.023	5.05%	19.81%

Table 6: 第 3 位委員 Success@K 統計摘要 (Bootstrap 95% CI, $n = 462$, 未命中率 70.56%)

K	Mean	95% CI	CI 寬度	Std	CV%	相對寬度%
10	0.006	[0.000, 0.015]	0.015	0.004	59.52%	233.33%
20	0.041	[0.024, 0.058]	0.035	0.009	21.52%	84.34%
30	0.080	[0.056, 0.106]	0.050	0.013	15.86%	62.16%
40	0.136	[0.106, 0.169]	0.063	0.016	11.74%	46.03%
50	0.186	[0.152, 0.223]	0.071	0.018	9.79%	38.37%
60	0.208	[0.171, 0.247]	0.076	0.019	9.30%	36.46%
80	0.262	[0.221, 0.301]	0.080	0.020	7.80%	30.58%
100	0.294	[0.253, 0.335]	0.082	0.021	7.13%	27.94%

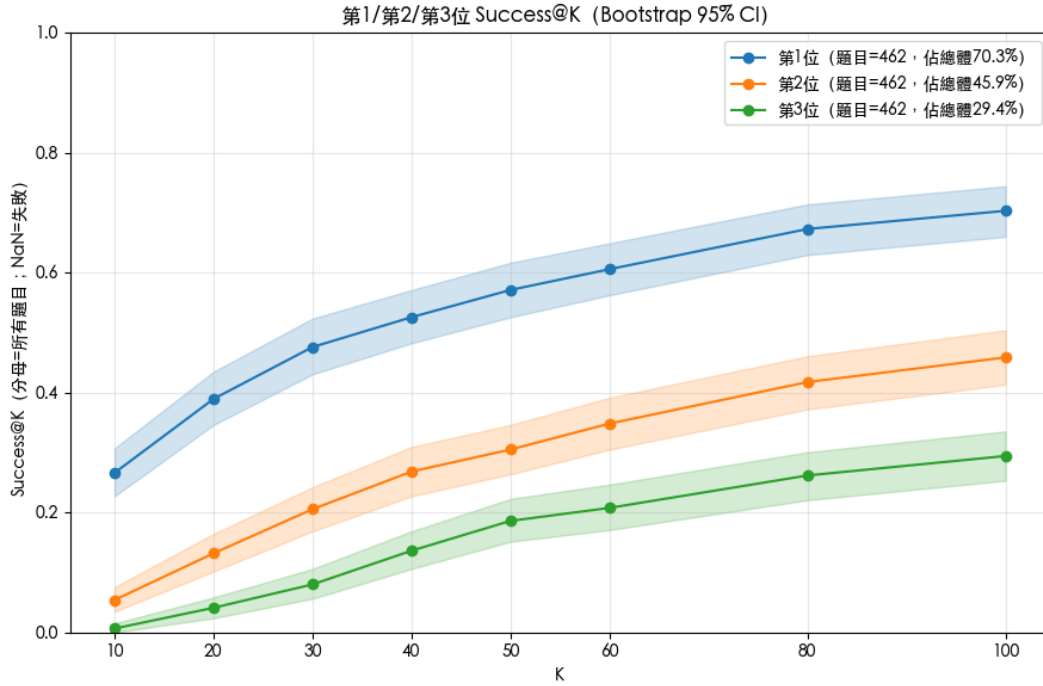


Figure 4: 前三位已知委員 Success@K 的 Bootstrap 95% 信賴區間

3.4 第一位命中位置分析

為進一步探討第一位已知委員的排序效果，我們針對成功命中的題目（共 325 題）分析其在推薦清單中的命中位置。分析項目包含平均位置、中位數、命中比例與 Bootstrap 95% 信賴區間，詳見表 7。

Table 7: 第一位委員命中位置與比例統計（僅計算成功命中題目）

統計指標	數值
平均命中位置（命中題目）	25.79
中位數命中位置（命中題目）	16.00
命中題數 / 總題數	325 / 462
命中比例	70.35%
完全未命中比例	29.65%
Bootstrap 95% 信賴區間	[23.06, 28.51]
CI 寬度	5.45
相對 CI 寬度 (CI/平均值)	21.13%
Bootstrap 樣本平均變異數	1.9487
變異係數 (CV)	5.41%

圖 5 為命中位置的直方圖，可觀察到大多數命中集中於推薦清單前 20 名以內，整體分布右偏。平均命中位置為 25.79，而使用 bootstrap 重抽樣 ($B = 1000$) 所計算的 95% 信賴區間為 [23.06, 28.51]，說明命中排序具穩定性。

圖 6 顯示相同資料的 Boxplot，其中中位數為第 16 名，進一步支持命中位置偏向前段。圖 7 為 ECDF 累積分布曲線，顯示約有 50% 的命中落在前 25 名內，證明系統能將高重要性委員推薦於前段名次。

綜合上述結果顯示，系統對第一位委員的排序效果良好，具有穩定性與統計意義，可有效提升推薦的實務應用性。其意義可說明至少一位重要在 25 名推薦名單內，或許能為後續實驗設計帶來參考價值。

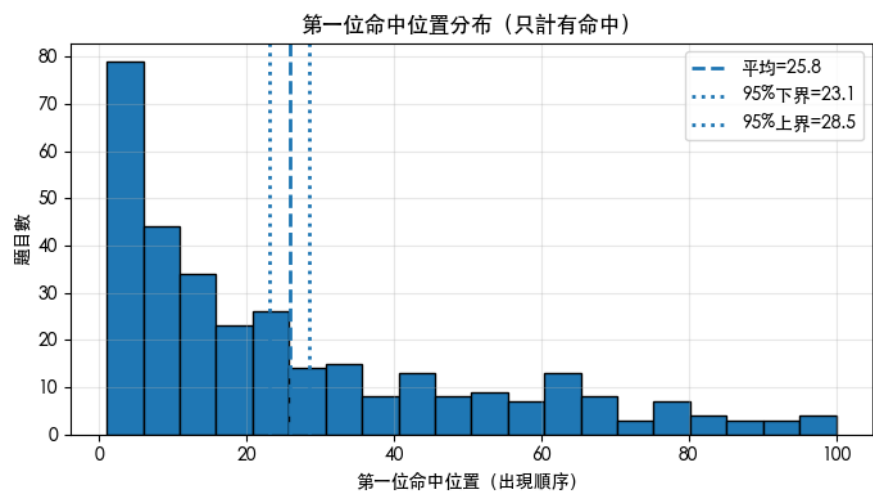


Figure 5: 第一位命中位置分布 (直方圖)

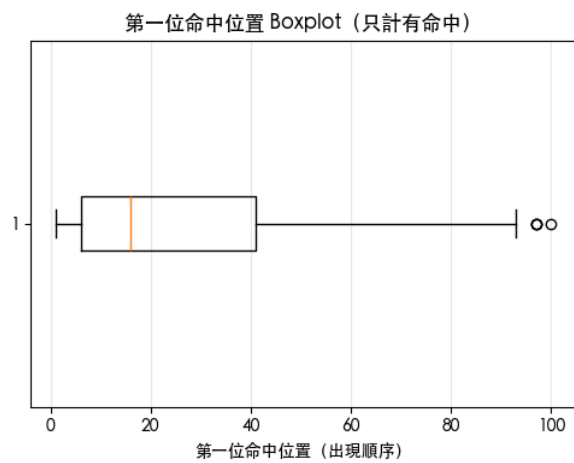


Figure 6: 第一位命中位置分布 (Boxplot)

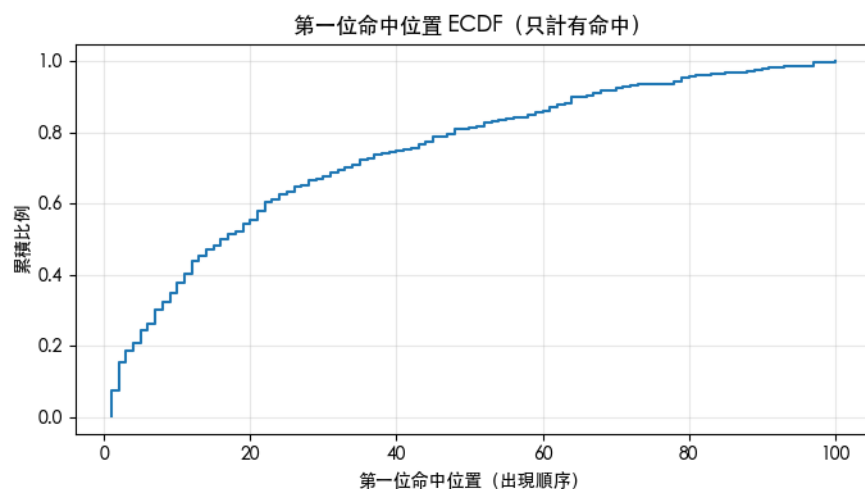


Figure 7: 第一位命中位置分布 (ECDF 累積曲線)

3.5 第二位命中位置分析

為進一步分析第二位已知委員的排序表現，我們針對成功命中的 212 題進行命中位置統計，包含平均位置、中位數、命中比例與 bootstrap 95% 信賴區間等指標，彙整於表 8。其中，平均命中位置為 40.24，中位數為第 36 名，顯示相較於第一位委員，排序明顯偏後。

Table 8: 第二位委員命中位置與比例統計 (僅計算成功命中題目)

統計指標	數值
平均命中位置 (命中題目)	40.24
中位數命中位置 (命中題目)	36.00
命中題數 / 總題數	212 / 462
命中比例	45.89%
完全未命中比例	54.11%
Bootstrap 平均命中位置	40.24
Bootstrap 95% 信賴區間	[36.66, 43.77]
CI 寬度	7.11
相對 CI 寬度 (CI/平均值)	17.67%
Bootstrap 樣本平均變異數	3.3151
變異係數 (CV)	4.52%

圖 8 為第二位命中的直方圖，可見命中位置分布較為分散，主要集中於第 30~60 名。Boxplot (圖 9) 呈現右尾離群值，對應平均位置較高的現象。ECDF (圖 10) 顯示約 50% 的命中落在推薦清單前 40 名內。

整體而言，第二位已知委員的命中比例約為 45.89%，其命中位置分布相較第一位更為分散，顯示在此位次下的排序不確定性較高。

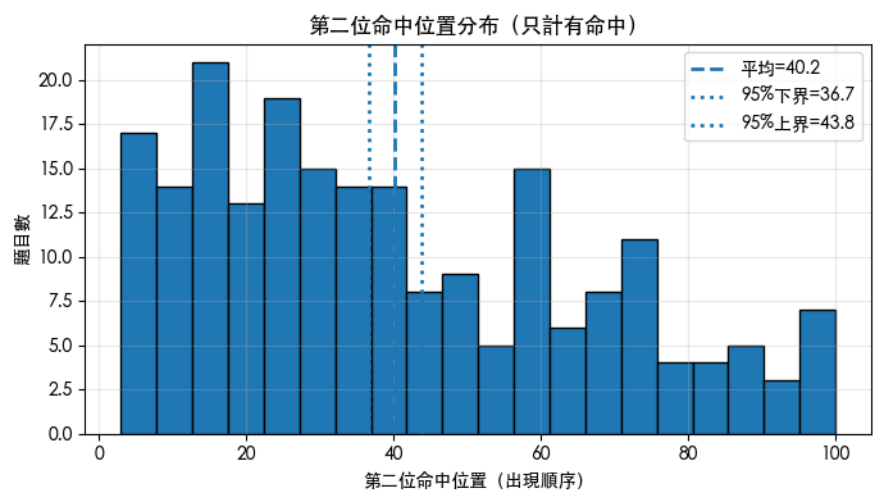


Figure 8: 第二位命中位置分布 (直方圖)

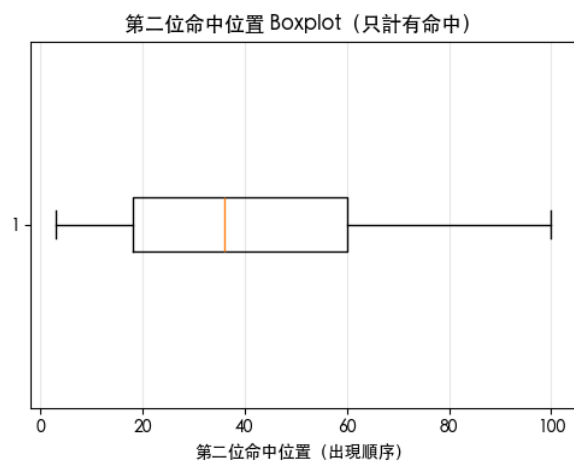


Figure 9: 第二位命中位置分布 (Boxplot)

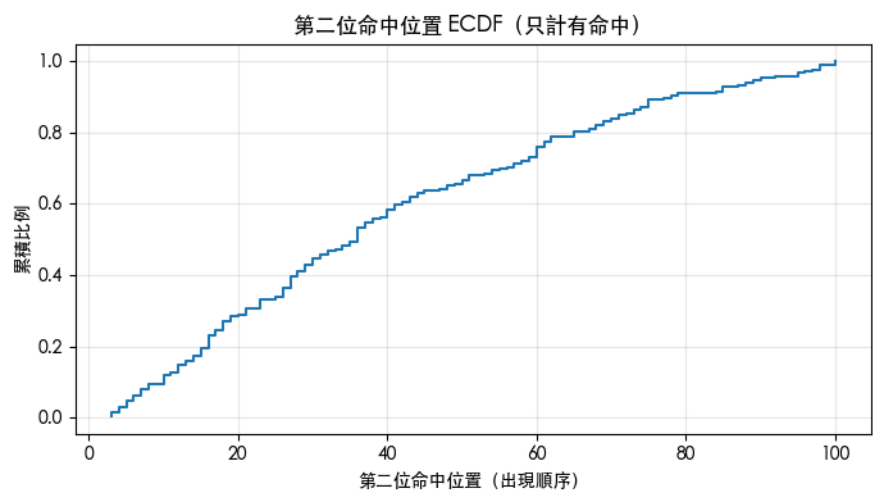


Figure 10: 第二位命中位置分布 (ECDF 累積曲線)

4 分領域表 (Per-Subject)

4.1 依學門的 Success@K 表現

圖 11 與表 9 顯示各學門在不同 K 下的 Success@K (第一位已知委員)。從結果中可觀察到以下幾點：

4.2 學門別重點 (精簡)

綜合表 9 與圖 11，重點如下：

- 表現較佳：政治、歷史、地理、法律、語言等學門在多數 K 值下 Success@K 均偏高，且題數充足，模型較易學到穩定訊號。
- 小樣本但亮眼：傳播學類僅 10 題，卻在中高 K 具高命中；惟樣本少。

綜合來看，系統在部分社會科學與語言相關學門上表現突出，而在題目數少或跨領域性強的學門上仍需改善。

Table 9: 各學門題數統計 (由多到少，橫向呈現)

學門	題數	學門	題數	學門	題數	學門	題數	學門	題數
政治學類	44	歷史學類	40	藝術學類	39	社會	34	法律學類	33
地理學類	32	心理學類	30	教育	27	人類學學類	25	未分類	24
語言學類	23	哲學類	20	經濟學類	20	文學	19	管理	17
外國語文學類	10	傳播學類	10	財務金融學類	8	體育學類	4	圖書	3
合計									462

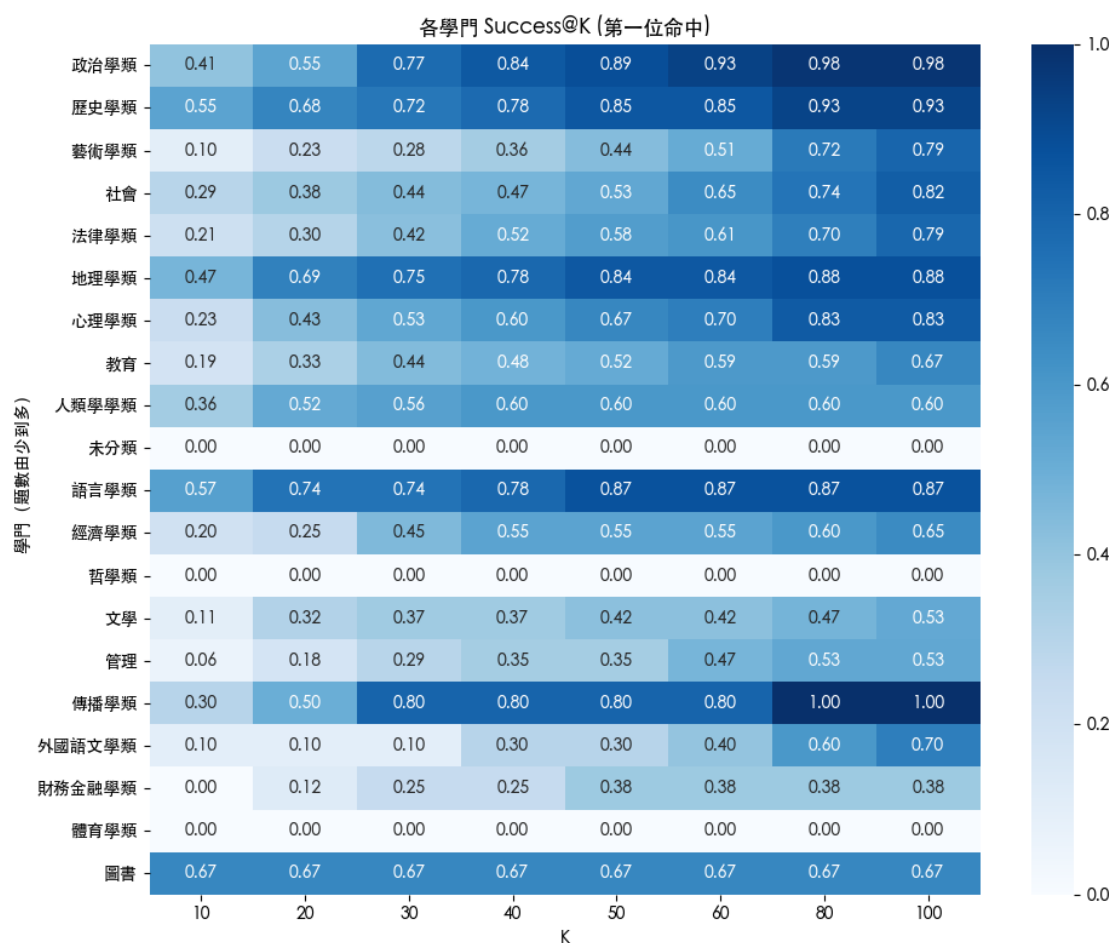


Figure 11: 各學門 Success@K 熱圖 (第一位命中)

4.3 各學門未命中第一位委員統計分析

為評估推薦系統在不同學門中的表現差異，我們統計各學門中「完全未命中第一位已知委員」的比例，作為排序系統可擴展性與泛化能力的觀察指標。統計指標包含各學門總題數、未命中題數與未命中比例，結果如表 10 和圖 12 所示。

Table 10: 各學門未命中第一位委員統計

學門	題數	未命中題數	未命中比例
未分類	24	24	1.000
哲學類	20	20	1.000
體育學類	4	4	1.000
財務金融學類	8	5	0.625
文學	19	9	0.474
管理	17	8	0.471
人類學學類	25	10	0.400
經濟學類	20	7	0.350
圖書	3	1	0.333
教育	27	9	0.333
外國語文學類	10	3	0.300
法律學類	33	7	0.212
藝術學類	39	8	0.205
社會	34	6	0.176
心理學類	30	5	0.167
語言學類	23	3	0.130
地理學類	32	4	0.125
歷史學類	40	3	0.075
政治學類	44	1	0.023
傳播學類	10	0	0.000

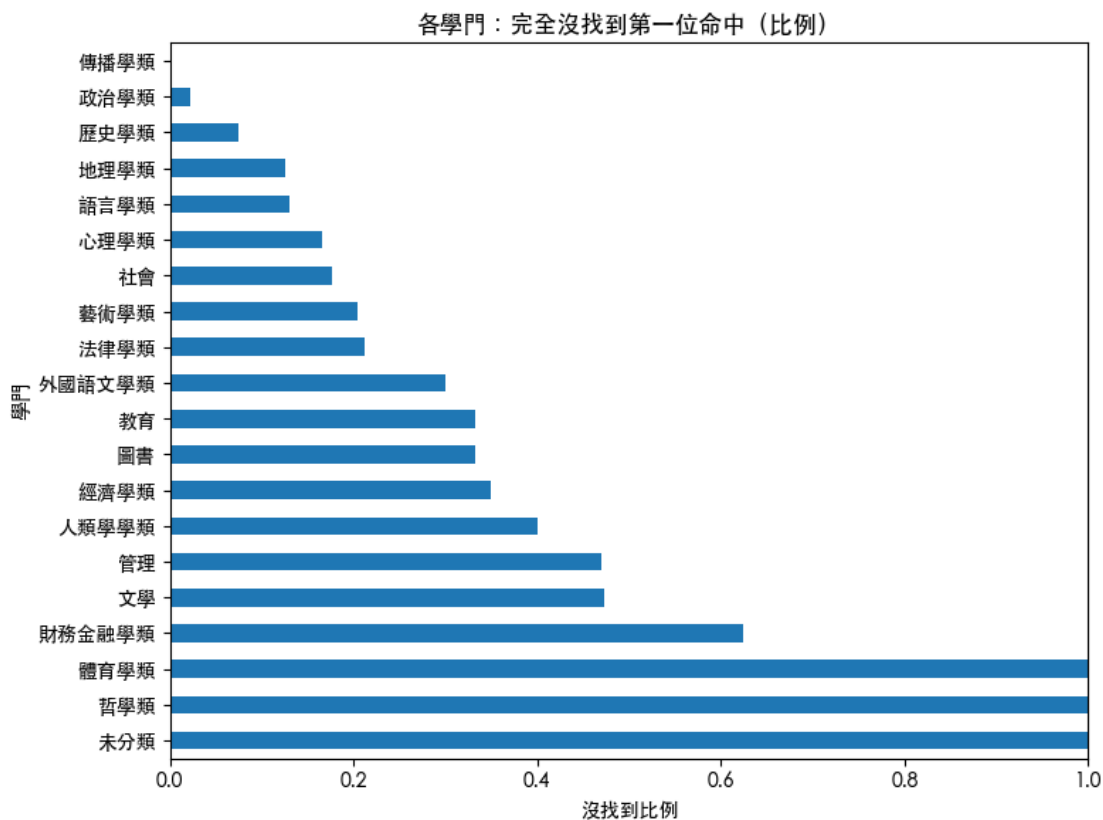


Figure 12: 各學門沒找到第一位命中（比例）

需要強調的是，本表所呈之未命中比例僅作為 ** 系統表現的觀察與描述性指標 **，** 不作為是否進行後續命中位置分析的依據 **。即便某學門未命中比例偏高，我們仍可能保留其在此其他統計與可

視化分析中的參與，以保留資料完整性與多樣性。

5 結論於討論

在後續的檢驗中，我們將所有題目的學門標籤統一改為「未分類」，觀察其對推薦結果的影響。實驗發現，系統在前 100 名內幾乎無法選到已知審查委員，故沒有審麼資訊，在此資料中不採用此方法

綜合上述結果顯示，系統對第一位委員的排序表現具有穩定性與統計意義，可在實務應用中有效提升推薦品質。其意義在於，即便在最差情況下，至少能確保一位關鍵委員落在前 25 名的推薦清單內，對實際審查召集人具有參考價值。

然而，在題目的學門資訊不明確或涉及跨領域時，系統表現比起有明確學門下較差，顯示單一領域導向的推薦方法仍有侷限。未來可考慮推薦名單的多樣性與公平性，以避免系統過度集中於特定學門或少數資深委員，進而提升系統在真實審查流程中的可行性。