

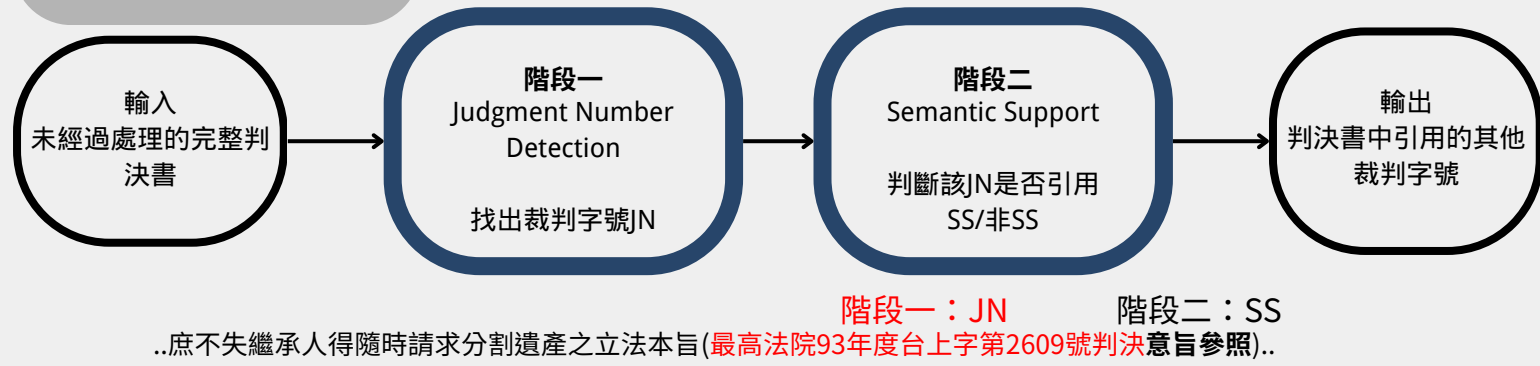
裁判字號引用偵測與分類

結合兩階段標註與資料自動化擴增策略
成員：陳冠任、陳恪行、徐少綸、薛溫塵

研究動機

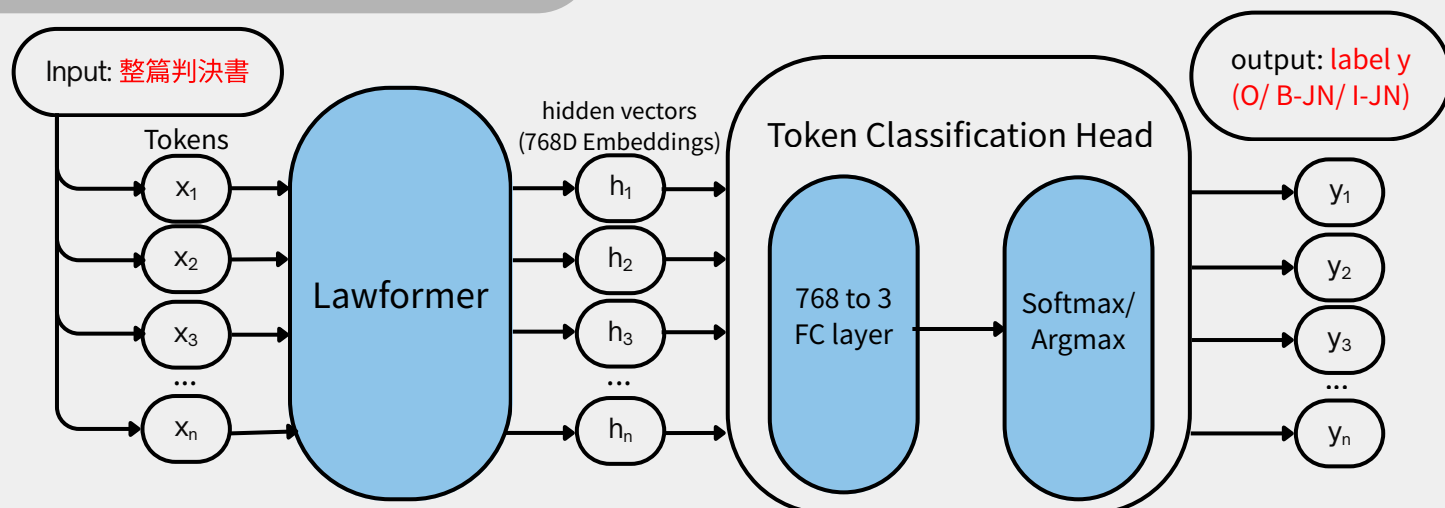
判決之間的引用關係，反映法院見解形成、延續與演變的脈絡。一份裁判引用哪些前案，往往代表其法律立場與論理來源；而一份裁判若被後續判決反覆引用，則可反映其在司法實務中的影響力與穩定性。因此，若能有效整理判決引用關係，便能協助使用者快速掌握關鍵前案、理解法律見解的發展方向，並提升法律檢索與研究效率。然而，現有法律資料庫多以全文檢索為主，使用者輸入關鍵字後，仍須自行從大量判決中判斷哪些裁判真正具有參考價值。由於判決書中常同時包含案件事實、程序經過、當事人主張與法院理由，裁判字號的出現並不一定代表法院實質引用該裁判作為論理基礎。因此，本研究希望從長篇判決書中自動辨識「被實質引用的裁判字號」，進一步協助建立判決之間的引用關係。

核心架構



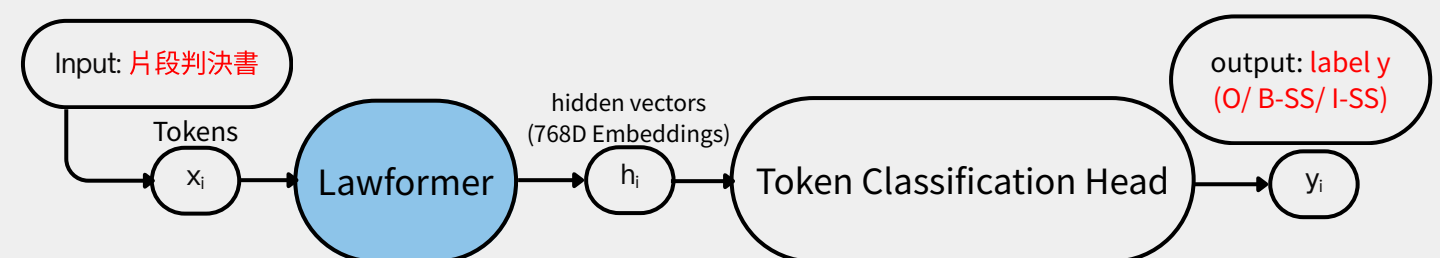
本研究在從長篇判決書中偵測被引用裁判字號時，主要面臨兩項挑戰。第一，裁判字號格式具有高度變異，若僅依賴正規表達式，規則過嚴容易漏判，規則過寬則容易誤判。第二，判決書中出現裁判字號並不必然代表實質引用，仍需依據上下文判斷其是法律見解上的引用，或只是案件事實與程序中的單純提及。因此，我們提出兩階段學習式方法：第一階段負責偵測判決書中的裁判字號 (Judgment Number, JN)，第二階段則根據 JN 附近的語意脈絡，判斷其是否構成實質引用，並辨識對應的引用關鍵字 (Semantic Support, SS)。

階段一：字號實體偵測



裁判字號 (JN) 通常是一段具有明確起點與終點的連續文字，為了從完整判決書中找出可能的裁判字號，因此本研究將此問題視為序列標註任務：採用 BIO 格式，B-JN 用來標示裁判字號的開頭，I-JN 用來標示裁判字號的內部，O 則代表其他非裁判字號文字，讓模型逐一判斷文本中的每個 token 是否屬於裁判字號。

階段二：引用關鍵字實體偵測



階段二與階段一同樣採用 Lawformer 加上 token classification head 的序列標註架構，但兩者的輸入與輸出標籤不同。階段二的輸入為經截斷後的短文本片段，也就是以單一 JN 為中心、保留其周邊語意脈絡即關鍵字出現分佈位置的片段。

階段一 + 階段二：引用字號實體偵測

最後會根據階段二的判斷結果，保留被判定為引用的 JN，並排除未構成引用的 JN。換言之，階段一提供「裁判字號的位置」，階段二提供「是否具有引用語意」的判斷，兩者整合後即可得到最終的「被引用裁判字號」標註結果。

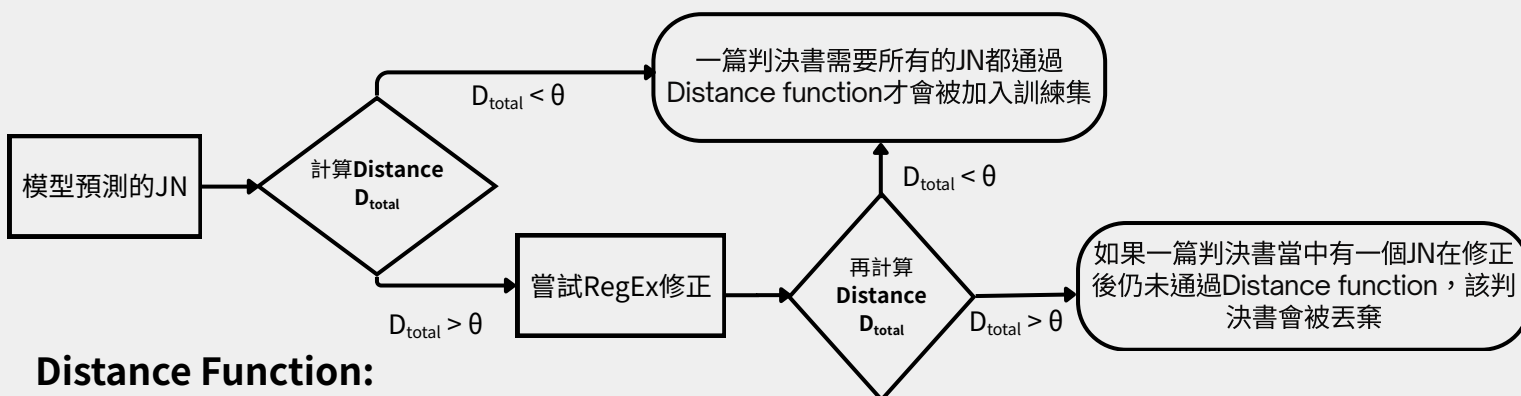
階段一：模型自我迭代

由於法律裁判書的標註需仰賴法律專業人士進行逐筆判讀，標註成本高昂且相當耗時。在標註資料有限的情況下容易造成模型泛化能力不足，影響其於實際應用中的有效性。因此，本研究提出一套自我迭代的新增資料方式，並同時引入一套基於 Distance Function 計算確保新增資料品質的方法：

P-all ensemble Upsample Pmodel Ensemble

此方法由「新增資料標註策略」+「資料篩選與修正流程」+「迭代對象」+「最終推論方式」四個部分組成
P-all ensemble Upsample Pmodel Ensemble

資料篩選與修正流程(Upsample)



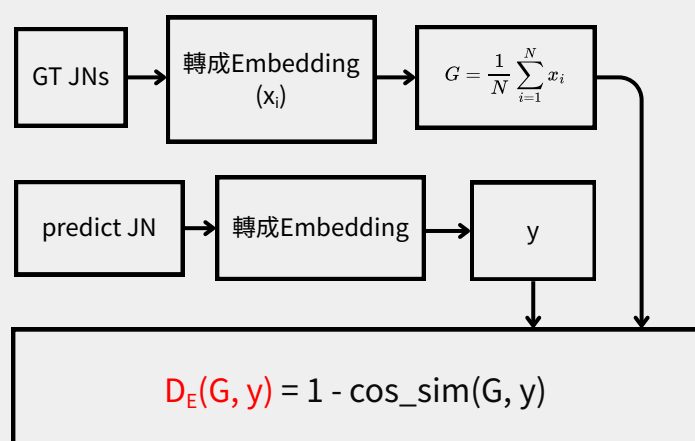
Distance Function:

$$D_{Total}(G, y) = \alpha * D_E(G, y) + (1 - \alpha) * D_R(y)$$

本研究提出之 Distance Function 結合裁判字號結構與語意特徵，分為兩部分：

$D_E(G, y)$ ：語意特徵(Embedding Distance)

用於衡量候選字號(y)與既有標註資料(G)於語意空間中的相似程度。



$D_R(y)$ ：結構特徵(Regex Distance)

用於衡量候選字號(y)與正確法律裁判字號格式規則之匹配程度。

$$D_R(y) = 0.7 * P_{ess} + 0.2 * P_{ct} + 0.1 * P_{len}$$

法院 + X年(度) + Y字 + (第)Z號 + 裁判類型

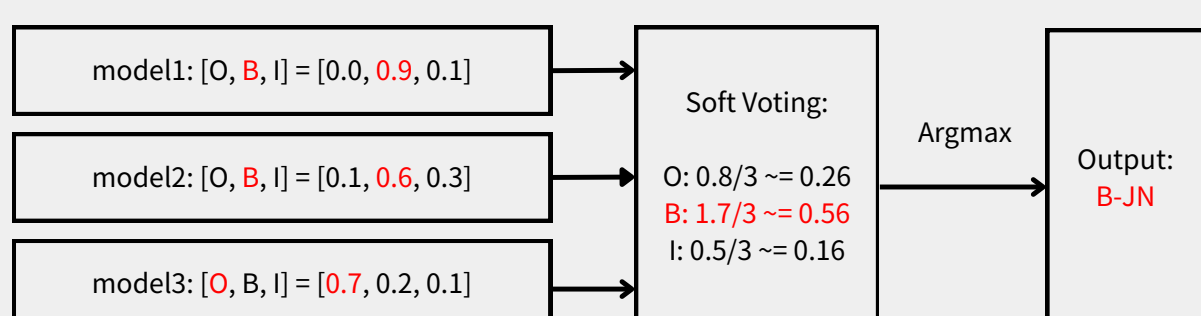
P_{ess} : 必要結構, 缺少時給予較高懲罰。

P_{ct} : 額外結構, 缺少時給予較低懲罰。

P_{len} : 偵測X, Y, Z字串長度是否異常偏離。

最終推論方式(Ensemble)

單一模型的穩定性相當有限。在進行標註的時候，如果某次的模型產生了錯誤，就會導致這類錯誤在後續的階段中持續累積，進而影響後續階段的訓練效果。Ensemble技術會讓多個模型同時參與該token的標籤機率決策，避免單一模型的失誤影響到整體的標註任務。

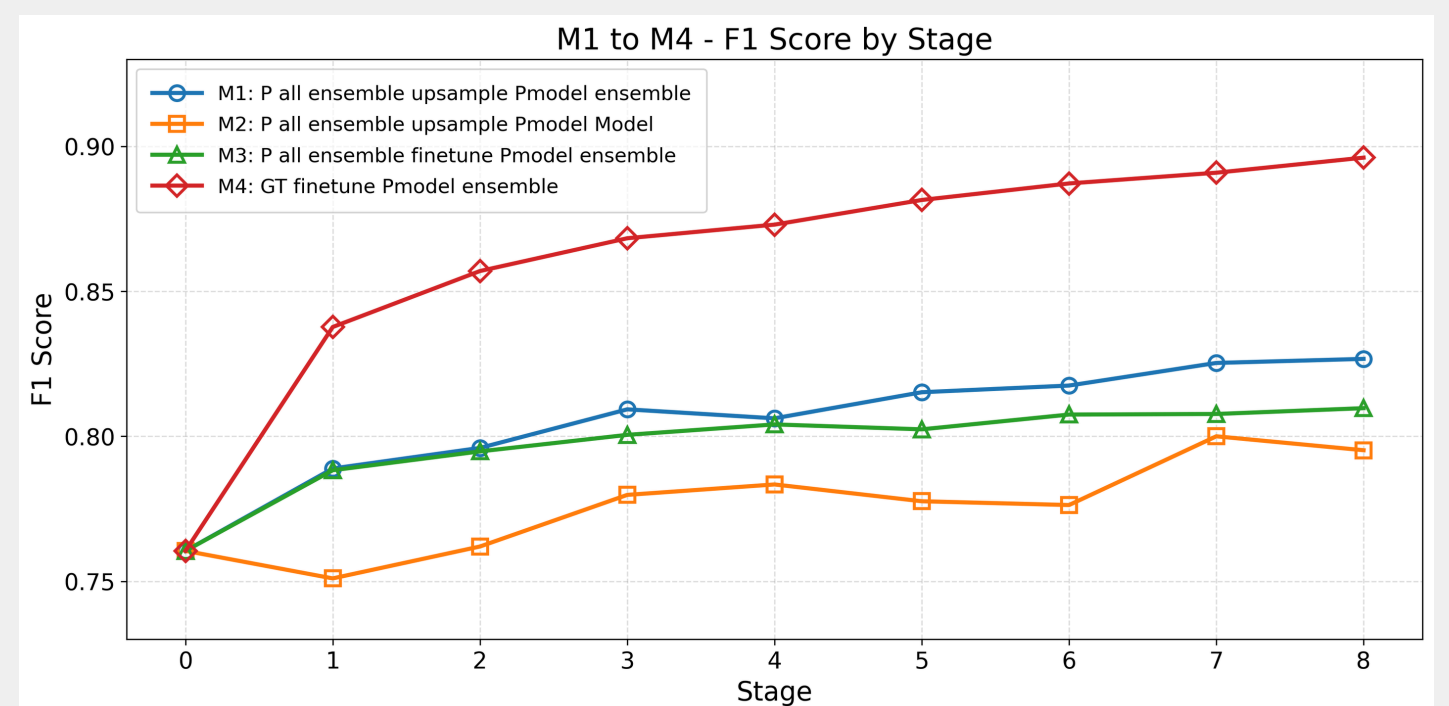


實驗結果

| 階段一方法 | Avg. Precision | Avg. Recall | Avg. F1 score |
|--------------|----------------|-------------|---------------|
| 階段一模型 | 0.8986 | 0.926 | 0.9123 |
| Regex | 0.7505 | 0.807 | 0.7776 |
| Regex Strict | 0.9565 | 0.1463 | 0.2535 |

| 階段二方法 | Avg. Precision | Avg. Recall | Avg. F1 score |
|----------|----------------|-------------|---------------|
| 階段二模型 | 0.9777 | 0.9787 | 0.9781 |
| Classify | 0.9566 | 0.9718 | 0.9641 |
| Keyword | 0.8868 | 0.9739 | 0.9282 |

針對方法P-all ensemble upsample Pmodel ensemble的「資料篩選與修正流程」和「最終推論方式」進行消融性實驗



| 階段一 + 階段二方法 | Avg. Precision | Avg. Recall | Avg. F1 score |
|-----------------|----------------|-------------|---------------|
| 分段式標註 | 0.9256 | 0.9193 | 0.9224 |
| Regex + Keyword | 0.755 | 0.58 | 0.655 |
| 一次標註是引用的JN | 0.8726 | 0.9029 | 0.8874 |

研究結論

- 本研究證明，語意理解模型能有效處理判決書中裁判字號格式多樣與引用語句表達不固定的問題，相較於單純依賴 Regex 或 keyword 的規則式方法，能取得更穩定且準確的辨識效果。
- 此外，在人工標註資料有限的情況下，透過 pseudo labeling、Distance Function 與信心值篩選進行模型自我迭代，具有實務可行性，能在控制資料品質的前提下逐步擴充訓練資料。
- 最後，實驗結果顯示，將任務拆分為 JN 裁判字號偵測與 SS 引用語意判斷兩個階段，比直接一次性標註被引用裁判字號更穩定，能有效提升判決引用關係辨識的整體表現。