

## 利用機器學習於中文法律文件之 標記、案件分類及量刑預測

### Exploiting Machine Learning Models for Chinese Legal Documents Labeling, Case Classification, and Sentencing Prediction

林琬真\*、郭宗廷\*、張桐嘉\*、顏厥安<sup>+</sup>、陳昭如<sup>+</sup>、林守德\*

Wan-Chen Lin, Tsung-Ting Kuo, Tung-Jia Chang,

Chueh-An Yen, Chao-Ju Chen and Shou-de Lin

#### 摘要

人工智慧於法學領域所發展出的法學資訊系統，主要提供參考資訊以協助司法審判；其重要議題包括法律文件分類、法律文件摘要、類似過去案例搜尋、協助判刑等。本論文著重探討「強盜罪」與「恐嚇取財罪」的分類，以及此兩種罪的量刑預測。我們針對「強盜罪」與「恐嚇取財罪」來定義 21 種法律要素標籤，並嘗試自動擷取所定義之標籤。實驗說明利用我們定義之法律要素標籤，能確實改善案件分類，以及進行量刑預測。最後，我們亦針對實驗結果，討論「強盜罪」與「恐嚇取財罪」的特徵，以及影響判刑長短之因素。

**關鍵字：**法律文件分類、量刑預測、強盜罪、恐嚇取財罪

---

\* 國立台灣大學資訊工程系 Department of Computer Science and Information Engineering, National Taiwan University

E-mail: myownstuff@hotmail.com; d97944007@csie.ntu.edu.tw; sdlin@csie.ntu.edu.tw

<sup>+</sup> 國立台灣大學法律學院 College of Law, National Taiwan University

E-mail: d97a21003@ntu.edu.tw; filawsof@ntu.edu.tw; tanchiauju@ntu.edu.tw

## Abstract

This paper exploits machine learning methods to separate robbery and intimidation cases, and predicting their sentencing by considering defined legal factors. We introduce a framework to fetch 21 legal factor labels of robbery and intimidation cases, then use the labels for case classification and sentencing prediction. Our experiments show that the legal factor labels can indeed improve the results of case classification and sentencing prediction. We then discuss the influence of these legal factors in both case classification and sentencing prediction tasks.

**Keywords:** Case Classification, Sentencing Prediction, Robbery, Intimidation

## 1. 緒論

在審判法律案件的法律體系大致可分為成文法(civil law)與判例法(common law)。判例法的特色在於遵循先例，法院在判決案件時依循相關判例來對目前審理的案件做判斷；成文法的特點在於有完整法典，法院依據成文法規作出判決，而判例做為參考不會拘束日後的判決。

台灣的法律體系屬於成文法，其中刑法明訂哪些不法行為屬犯罪行為，以及對於這些犯罪行為應如何科處刑罰；但在法官判決，先前之判例仍具有一定的參考價值。然而，儘管法律條文已明確列出各種犯罪行為，在於實際判斷上仍具有模糊地帶。例如刑法中「強盜罪」與「恐嚇取財罪」具有類似不法構成要素：刑法第 328 條第 1 項定義普通強盜罪：「意圖為自己或第三人不法之所有，以強暴、脅迫、藥劑、催眠術或他法，至使不能抗拒，而取他人之物或使其交付者，為強盜罪，處五年以上有期徒刑」；刑法第 346 條第 1 項定義恐嚇取財罪：「意圖為自己或第三人不法之所有，以恐嚇使人將本人或第三人之物交付者，處六月以上五年以下有期徒刑，得併科一千元以下罰金」。兩者之差異主要在於強盜罪的行為人犯罪行為脅迫程度，足以使得被害人不能抗拒。然而此差異在實際案例上，卻容易造成判斷混淆，例如「行為人持槍進入超商叫被害人把錢拿出來」以及「行為人持美工刀進入超商叫被害人把錢拿出來」，前者判屬強盜罪而後者則屬恐嚇取財罪。另一方面，兩罪的刑期相差甚大（前者是五年以上，後者是五年以下），對於嫌疑人而言影響亦甚鉅。因此，一個能支援及協助法官判別「易有模糊地帶之相關罪行」，乃至進一步提供建議刑期的系統，便極為重要。

但是，要建置這樣的系統，會面臨數個挑戰。首先，此系統需能自動標記法律要素標籤，而不需額外之人工參與。其次，此系統需以法律要素標籤自動進行案件分類及量刑預測。最後，系統所做出的建議及預測結果，亦需經由仔細的檢查和討論，以確保其可信度。

為了解決上述的問題，本研究針對「強盜罪」與「恐嚇取財罪」定義法律要素，並期望達成自動標記法律要素，接著利用法律要素資訊來分類「強盜罪」與「恐嚇取財罪」以及預測此兩種罪的判處刑期，最後討論「強盜罪」與「恐嚇取財罪」的分類特徵以及

影響判刑的因素。

本文之內容安排如下。在第二節，我們將探討及比較相關文獻。在第三節，我們將介紹自動標記方法及實驗結果。在第四節及第五節，我們以人工標記分別說明案件分類及量刑預測之方法和結果。在第六節，我們會結合自動標記來進行案件分類及量刑預測。最後在第七節，我們將提出結論及未來之可能方向。

## 2. 相關研究

人工智慧結合法律領域的法資訊學在國外已研究多年，目的在於增進司法審判的效率以及協助撰寫法律文件，主要研究包括法律文件分類、法律文件總結、相關案件檢索、協助判決等。

在於法律文件分類研究上，Ashley (Bru'ninghaus, 2009)等人提出 SMILE 系統處理營業秘密法案件。依據事先定義之要素，以人工標記方式建立案例資料庫，並使用決策樹演算法來分類案件。Lame (Lame, 2001)等人針對法國法律文件是先建立常用字表，以 Term frequency-inverse document frequency (TF-IDF)為權重並使用 Support vector machine (SVM)進行法律文件分類。在處理中文法律文件分類研究中，Liu (LIU, 2004)等人建立一個類 Case based reasoning (CBR)系統來分類 12 種罪，此系統依據事先定義的各種犯罪規則建立案例資料庫，並使用 k-nearest neighbor algorithm (kNN)選出與處理案件最相似之案例以達分類目的。在我們的研究中，除利用詞彙的 TF-IDF 資訊外，尚針對要分類的犯罪設計法律要素標籤以提供犯罪資訊，希望能利用 TF-IDF 以外的資訊幫助分類案件以及預測量刑。

在量刑預測研究上，Schild (SCHILD, 1998)提出一個 CBR 系統來預測強盜案件以及強制性交案件的量刑，系統詢問使用者事先定義之相關問題，根據使用者的回答利用決策樹演算法找到類似案例，最後依據過往案例來預測量刑。在處理中文法律案件量刑方面，司法院於 2011 年提出量刑資訊系統(Kuo *et al.*, 2006)，針對妨害性自主罪設計量刑因子，對於量刑因子資訊進行統計分析，提供妨害性自主罪整體量刑分布、分析量刑加重或酌減之原因。

而我們研究中則是提出的法律要素標籤資訊，並且希望以自動標記來取代人工標記。在於連續標記方面 Wei Jiang 等人(Jiang, 2006)比較 HMM、MEMM、CRF 以及 SVM 在於連續標記詞性上的效果，實驗結果顯示 CRF 在於連續標記任務上較其他模型有良好表現。另外本研究也參考 Zhang Chengmin 等人(Zhang, 2008)探討 CRF 在於標記任務上使用的特徵。

## 3. 自動標記法律要素標籤

每種犯罪行為有各自的構成要素包括行為主體、行為客體、行為、行為時之特別情狀等等，不法行為必須符合特定的「構成要件要素」，並且成立「違法性」與「有責性」才屬犯罪行為，而本研究中單純就「構成要件要素」來對法律案件做分類及量刑預測。我

們針對「強盜罪」與「恐嚇取財罪」定義 21 種法律要素標籤，期望能表達法律案件中的犯罪構成要素。標籤依性質分為兩大類：

- **Global labels**：法官認定事實、量刑、檢察官起訴內容、被告辯護、證人證詞，共 5 種。
- **Local labels**：行為人、犯罪行為、被害人、被害人反應、財物轉移態樣、行為人主觀要素、犯案時間、犯案工具、犯案地點、行為人特徵、被害人特徵、犯後態度、行為人動機、行為人與被害人關係、財物類型、共犯，共 16 種。

我們由植根法律網(植根法律網)收集 2006 年到 2010 年總共 21 個地方法院的刑事判決書，去除裁定、簡易判決、上訴駁回、不起訴、無罪等案件後，總計強盜與恐嚇取財案件共 2113 件。考量到人工標記所需之時間及人力成本，我們隨機選取了 140 件進行人工標記，案件分布如表 1，並使用中研院斷詞系統(Hotho *et al.*, 2003)對語料進行斷詞。

表 1. 人工標記判決書分布情形

| 地方法院 | 台北 | 士林 | 板橋 | 宜蘭 | 基隆 | 桃園 | 新竹 | 彰化 | 南投 | 台南 | 花蓮 | 台東 | 屏東 | 總件數 |
|------|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|
| 恐嚇取財 | 0  | 2  | 17 | 11 | 1  | 6  | 9  | 3  | 2  | 9  | 4  | 3  | 0  | 67  |
| 強盜   | 1  | 5  | 13 | 6  | 1  | 4  | 4  | 6  | 7  | 8  | 7  | 2  | 9  | 73  |

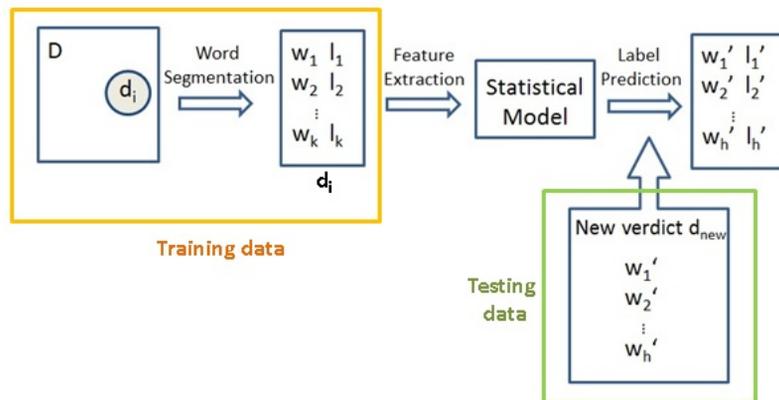


圖 1. 自動標記架構

我們定義  $D$  為判決書之集合  $D = \{d_i | i = 1, 2, \dots, m\}$ ， $L$  為定義標籤集合  $L = \{l_j | j = 1, 2, \dots, n, n = \text{number of labels}\}$ ，每篇判決書  $d_i$  經過斷詞處理後，得到  $W_i$  為判決書  $d_i$  之斷詞集合，其中  $W_i = \{w_k | k \in N\}$ 。而自動標記問題定義如下：給定一篇斷詞形式的判決書  $d_{new} = W_{new}$ ，判斷  $W_{new}$  中的每個斷詞  $w_k$  所應標記的標籤  $l_j$ ，自動標記架構如圖 1 所示。我們運用機率模型來進行自動標記，在給定  $x$  為一含有  $n$  個字的序列， $y$  為

對應之標籤序列，機率模型中的 Conditional random fields (CRF) 模型表示式如下：

$$p(y|x) = \frac{\exp\sum_j \lambda_j F_j(x,y)}{Z(x,\lambda)}$$

$$Z(x,\lambda) = \sum_y \exp\sum_j \lambda_j F_j(x,y)$$

$$F_j(x,y) = \sum_i f_j(y_{i-1}, y_i, x, i), \text{ 其中 } i \text{ 表在序列 } x \text{ 中的位置}$$

由於 CRF 模型中特徵函式能考量位於序列的位置，因此能考慮長距離上下文資訊，在於連續標記有極好的效能，故使用 CRF 模型實作法律要素標籤自動標記。

### 3.1 特徵值處理

由於定義的標籤性質不同，我們先標記 Global 再標記 Local labels。使用特徵為：

- 斷詞以及詞性：由中研院斷詞系統(Hotho *et al.*, 2003)取得斷詞及詞性資訊。
- 斷詞位於判決書中的位置：以「：」、「；」以及「。」判斷句子，並進一步計算斷詞所在的句子在於整篇判決書中的位置。在實驗中，我們分別將整份判決書分為三等份、四等份、五等份以及十等份。以三等份為例，每個斷詞的位置資訊即代表此斷詞位於一篇判決書的前 1/3 或中 1/3 或者後 1/3。
- Cosine-normalized tfidf：對於每種標籤  $k$ ，計算動詞及名詞之特徵如下：

$tf_{i,j}$  = 在文件  $j$  中，動詞或名詞  $i$  被標為標籤  $k$  的次數

$df_i$  = 多少篇判決書中有動詞或名詞  $i$  被標為標籤  $k$

$$idf_i = \log\left(\frac{|D|}{df_i}\right)$$

$$tfidf_{i,j} = tf_{i,j} * idf_i$$

$$CosNorm(tfidf_{i,j}) = \frac{tfidf_{i,j}}{\sqrt{\sum_{k \in all \ labels} tfidf_{i,j}^2}}$$

標記 Local labels 時，除了以上三種特徵外，亦加入以我們系統所預測的 Global labels 值作為特徵。

### 3.2 實驗結果

我們使用 CRF++(Wibowo & Williams, 2002)實作自動標記，取五疊交叉驗證(5-fold cross validation)以 F1 score 做評估，實驗結果如表 2 及表 3 所示。由實驗結果我們發現，由於 global label 通常標記一大段文字，所以判決書位置分得細標記結果反而不好；相對地當判決書位置分得越細 local label 標記結果越好。

表 2. 自動標記 Global label 結果

| 特徵   | 恐嚇取財         | 強盜           |
|--|--------------|--------------|
| 斷詞、詞性                                      | 0.577        | 0.515        |
| 斷詞、詞性、位置 1/3                               | 0.626        | 0.557        |
| 斷詞、詞性、位置 1/4                               | 0.624        | 0.555        |
| 斷詞、詞性、位置 1/5                               | 0.619        | 0.546        |
| 斷詞、詞性、位置 1/10                              | 0.595        | 0.542        |
| 斷詞、詞性、動詞名詞 Cosine-normalized TF-IDF        | 0.741        | 0.782        |
| 斷詞、詞性、位置 1/3、動詞名詞 Cosine-normalized TF-IDF | <b>0.818</b> | <b>0.834</b> |

表 3. 自動標記 Local label 結果

| 特徵  | 恐嚇取財         | 強盜           |
|---|--------------|--------------|
| 斷詞、詞性   | 0.350        | 0.307        |
| 斷詞、詞性、位置 1/3  | 0.366        | 0.323        |
| 斷詞、詞性、位置 1/4  | 0.375        | 0.328        |
| 斷詞、詞性、位置 1/5  | 0.396        | 0.353        |
| 斷詞、詞性、位置 1/10   | 0.403        | 0.361        |
| 斷詞、詞性、預測的 Global label 資訊   | 0.352        | 0.310        |
| 斷詞、詞性、動詞名詞 Cosine-normalized TF-IDF                                 | 0.540        | 0.562        |
| 斷詞、詞性、位置 1/10、動詞名詞 Cosine-normalized TF-IDF、<br>預測的 Global label 資訊 | <b>0.604</b> | <b>0.615</b> |

### 4. 以人工標籤進行案件分類

在強盜罪與恐嚇取財罪的分類上，「行為的脅迫程度是否足以令被害人不能抗拒」是區分兩種案件的關鍵。因此，我們利用法律要素標籤來改善分類效果，並探討影響分類結果的法律要素標籤資訊。另外，由於判決是為法官對於某犯罪的判決紀錄，因此判決書

中的某些部份（例如「案由」）可能會提及案件所屬犯罪類型以及所判刑罰。故此，我們只取用「法官認定事實」部份作為輸入，以求輸入資料不涉及告知犯罪類型以及判處的量刑，避免看到答案。我們首先用直覺的規則來區分強盜案件以及恐嚇取財案件：當案件事實中出現「不能抗拒」則分類此案件為強盜案件；當案件事實中出現「心生畏懼」且無出現「不能抗拒」則分類此案件為恐嚇取財案件，同時也考慮關鍵字前是否出現「未致」、「未達」、「不造成」等否定字。以此規則判斷分類的結果 F1 score 為 0.852；我們以此結果做為基準線，希望加上標籤資訊後可以有所改進。

#### 4.1 使用特徵與實驗結果

我們使用法律要素標籤資訊以進行案件分類，使用特徵包括：

- Fact\_tfidf：對於被標記為「法官認定事實」的斷詞，計算 TF-IDF 值。
- Local label 標記次數：各 local label 被標記的次數；原始標記資料中，若有兩個斷詞標同一種 label，只以標一次計算。
- Local label 標記順序：各 local label 的標記順序值。

我們使用四疊交叉驗證(4-fold cross validation)以 F1 score 做評估，並比較兩個分類器 Liblinear 與 Logistic model tree (LMT) 的分類結果如表 4 及表 5。結果顯示使用 Fact\_tfidf，local label 標記次數，和 Local label 標記順序為特徵，並採 LMT 做分類之結果最佳，可達 F1 score 為 0.943。另外，由 Fact\_tfidf 中篩選出分類重要關鍵字如表 6。若濾掉這些關鍵字後，同樣使用 Fact\_tfidf，Local label 標記次數，和 Local label 標記順序為特徵，並採 LMT 做分類的 F1 score 降為 0.721。因此，我們相信這些關鍵字對於案件分類是相當重要的。

表 4. Liblinear 分類結果

| 特徵                         | 特徵數   | F1 score     |
|----------------------------|-------|--------------|
| Fact_tfidf (Ftfidf)        | 10034 | <b>0.834</b> |
| Local label frequency (LF) | 16    | 0.615        |
| Local label order (LO)     | 256   | 0.695        |
| Ftfidf + LF                | 10050 | 0.702        |
| Ftfidf + LO                | 10290 | 0.769        |
| LF + LO                    | 272   | 0.642        |
| Ftfidf + LF + LO           | 10306 | 0.698        |

表 5. LMT 分類結果

| 特徵                         | 特徵數   | F1 score     |
|----------------------------|-------|--------------|
| Fact_tfidf (Ftfidf)        | 10034 | 0.929        |
| Local label frequency (LF) | 16    | 0.654        |
| Local label order (LO)     | 256   | 0.693        |
| Ftfidf + LF                | 10050 | 0.936        |
| Ftfidf + LO                | 10290 | 0.921        |
| LF + LO                    | 272   | 0.678        |
| Ftfidf + LF + LO           | 10306 | <b>0.943</b> |

表 6. 分類關鍵字

| 恐嚇取財                         | 強盜  |
|------------------------------|---|
| 心生畏懼、恐嚇、取財、交付、晚間、拿起、意圖、取走、受有 | 強盜、抗拒、不能、脅迫、客觀、強暴、強取、現金、以致、得手、現場、身體、足以、強行 |

因此，我們選擇上述 Fact\_tfidf，重要關鍵字，及其他 Local label 資訊，最後採用的特徵如表 7。為求進一步驗證及提升結果，我們使用單一特徵、所有特徵、Leave one out (LOO)以及以 Forward selection 以找出最佳特徵組合，其實驗結果如表 8、9 及 10。從表 10 可知，最佳組合之 F1 score 為 0.957。

表 7. 分類使用特徵

| 特徵                      | 特徵描述                         | 特徵數   |
|-------------------------|------------------------------|-------|
| Fact_tfidf (ftfidf)     | 標記為「事實」區段文字之 TF-IDF 值        | 10034 |
| Behavior_tfidf (btfidf) | 標記為「行為」區段文字之 TF-IDF 值        | 2897  |
| 行為關鍵字(bkey)             | 標記為「行為」的文字中是否出現表五所列之關鍵字      | 23    |
| 被害人反應關鍵字 (vkey)         | 標記為「被害人反應」的文字中是否出現表五所列之關鍵字   | 23    |
| 行為人個數(actor)            | 1~2 人、3~5 人、6 人以上            | 3     |
| 行為人特徵(afeature)         | 行為人犯罪背景描述，包括無前科、有前科、有精神疾病、累犯 | 4     |
| 共犯個數(accomplice)        | 0 共犯、1~2 人、3~5 人、6 人以上       | 4     |
| 被害人個數(victim)           | 1~2 人、3~5 人、6 人以上            | 3     |
| 犯後態度(after)             | 未坦承犯行/否認犯行/未見悔意/態度不佳、坦承部分    | 3     |

|                  |   |     |
|------------------|---|-----|
|                  | 犯行、坦承犯行   |     |
| 財物(property)     | 分爲一千以下、一千到五千、五千到一萬、一萬到十萬、十萬到百萬、錢、百萬以上、手機電子用品、本票等文件、證件、交通工具、珠寶金飾、皮包、其他 | 14  |
| 犯案工具(tool)       | 分爲電話、錄影光碟、廣義刀械、槍、道具槍、言語行爲、交通工具、棒棍、電擊棒、縱火工具、信函、掩飾衣物、網綁工具、衣物、藥物、金屬工具、其他 | 17  |
| Local label 標記次數 | 各 local label 在判決書中被標記的次數，原始標記資料中若兩斷詞標同一種 label 算標一次                  | 16  |
| Local label 標記順序 | 各 local label 在判決書中標記的順序  | 256 |

表8. 使用單一及所有特徵實驗結果

| 特徵               | 特徵數   | F1 score |
|------------------|-------|----------|
| Fact_tfidf       | 10034 | 0.929    |
| Local label 標記次數 | 16    | 0.654    |
| Local label 標記順序 | 256   | 0.693    |
| 行爲人個數            | 3     | 0.374    |
| 行爲人特徵            | 4     | 0.584    |
| 犯後態度             | 3     | 0.557    |
| Behavior_tfidf   | 2897  | 0.785    |
| 行爲關鍵字            | 23    | 0.779    |
| 被害人犯影關鍵字         | 23    | 0.751    |
| 被害人個數            | 3     | 0.367    |
| 共犯個數             | 4     | 0.512    |
| 財物               | 14    | 0.603    |
| 犯案工具             | 17    | 0.670    |

表9. LOO 實驗結果

| 特徵               | 特徵數   | F1 score |
|------------------|-------|----------|
| Fact_tfidf       | 3263  | 0.850    |
| Local label 標記次數 | 13281 | 0.921    |
| Local label 標記順序 | 13041 | 0.936    |
| 行為人個數            | 13294 | 0.936    |
| 行為人特徵            | 13293 | 0.936    |
| 犯後態度             | 13294 | 0.936    |
| Behavior_tfidf   | 10400 | 0.929    |
| 行為關鍵字            | 13274 | 0.957    |
| 被害人反應關鍵字         | 13274 | 0.943    |
| 被害人個數            | 13294 | 0.929    |
| 共犯個數             | 13293 | 0.943    |
| 財物               | 13283 | 0.936    |
| 犯案工具             | 13280 | 0.943    |

表10. 所有特徵及 Forward selection 實驗結果

| 特徵  | 特徵數   | F1 score     |
|---|-------|--------------|
| 所有特徵  | 13297 | 0.936        |
| Fact_tfidf、Behavior_tfidf、被害人反應關鍵字、Local label 標記次數 | 12970 | <b>0.957</b> |

## 4.2 討論

我們希望探討每個特徵對於分類的影響，因此依據特徵值的權重來進行分析。特徵值的權重如表 11。由於前 20 項特徵絕大多數為 TF-IDF 特徵，為了解「非 TF-IDF 特徵」對於分類的影響，在表 12 中，我們列出分類強盜罪及恐嚇取財罪中，「非 TF-IDF 特徵」的權重。我們對於「非 TF-IDF 特徵值」做討論，並對於人工標記語料作統計佐以驗證如表 13。

表 11. 分類恐嚇取財與強盜的前 10 項特徵值權重

| 強盜    |           | 恐嚇取財  |                            |
|-------|-----------|-------|----------------------------|
| 權重    | 特徵        | 權重    | 特徵                         |
| 16.2  | ftfidf_不能 | -5.22 | ftfidf_恐嚇                  |
| 15.41 | ftfidf_抗拒 | -2.84 | bkey_恐嚇                    |
| 12.93 | ftfidf_徒手 | -2.71 | vkey_心生畏懼                  |
| 10.7  | ftfidf_旋即 | -2.68 | ftfidf_恐嚇                  |
| 9.56  | ftfidf_可信 | -2.63 | ftfidf_心生畏懼                |
| 9.26  | ftfidf_有無 | -2.58 | ftfidf_否則                  |
| 7.71  | btfidf_抵住 | -2.28 | btfidf_恫稱                  |
| 6.81  | vkey_抗拒   | -2.13 | after_未坦承犯行/否認犯行/未見悔意/態度不佳 |
| 5.47  | ftfidf_強盜 | -2.05 | ftfidf_聯絡                  |
| 5.4   | ftfidf_得手 | -2.01 | btfidf_恐嚇                  |

表 12. 分類強盜罪及恐嚇取財罪非 TF-IDF 前 10 項特徵的權重

| 強盜   |               | 恐嚇取財  |             |
|------|---------------|-------|-------------|
| 權重   | 特徵            | 權重    | 特徵          |
| 6.81 | vkey_抗拒       | -2.84 | bkey_恐嚇     |
| 4.54 | bkey_抗拒       | -2.71 | vkey_心生畏懼   |
| 4.44 | tool_槍        | -2.13 | after_未坦承犯行 |
| 3.42 | bkey_強行       | -1.57 | bkey_心生畏懼   |
| 2.85 | tool_刀械       | -1.31 | bkey_拿起     |
| 2.33 | tool_道具槍      | -0.68 | tool_電話     |
| 1.88 | bkey_強取       | -0.62 | bkey_撥打     |
| 1.86 | bkey_取走       | -0.6  | vkey_交付     |
| 1.85 | property_珠寶金飾 | -0.59 | vkey_以致     |
| 1.82 | property_皮包   | -0.57 | tool_言語行爲   |

表 13. 分類特徵討論

| 特徵       | 強盜  | 恐嚇取財                     |
|----------|---|--------------------------|
| 行為關鍵字    | 抗拒、強行、強取、取走、強暴、取財、得手、客觀、意圖、不能               | 恐嚇、心生畏懼、拿起、撥打            |
|          | 使用上列關鍵字的案件中有 72.3% 為強盜案件                    | 使用上列關鍵字的案件中有 90.2% 為恐嚇取財 |
| 被害人反應關鍵字 | 抗拒、不能                                       | 心生畏懼、交付、以致               |
|          | 使用上列關鍵字的案件中有 93.8% 為強盜案件                    | 使用上列關鍵字的案件中有 80.4% 為恐嚇取財 |
| 犯案工具     | 槍、刀械、道具槍、金屬工具(ex 鐵鋸)、換裝衣物                   | 電話、言語行為、信函               |
|          | 使用上列工具的案件中有 79.6% 為強盜案件                     | 使用上列工具的案件中有 85.7% 為恐嚇取財  |
| 財物       | 珠寶金飾、皮包、證件、其他(ex 保險箱 鑰匙)、金額一千以下、一千到五千、一萬到十萬 | 本票、金額十萬到百萬、百萬以上          |
|          | 含有上列財物的案件中有 76.6% 為強盜案件                     | 含有上列財物的案件中有 77.3% 為恐嚇取財  |
| 犯後態度     | 坦承部分犯行                                      | 未坦承犯行/態度不佳、坦承犯行          |
|          | 有上列特徵的案件中有 75% 為強盜案件                        | 有上列特徵的案件中有 53.4% 為恐嚇取財   |
| 行為人特徵    | 精神疾病、無前科                                    |                          |
|          | 有上列特徵的案件中有 61.5% 為強盜案件                      |                          |

## 5. 以人工標籤進行量刑預測

在案件分類後，我們希望可以進一步預測強盜案件以及恐嚇取財案件的有期徒刑刑期。訓練語料與案件分類時相同。另外，考量到「一個案件具有多個行為人，而每個行為人的刑期可能會不相同」的情形，我們將每個行為人獨立成新案件；因此，訓練語料總案件數增為 155 件。強盜罪在人工標記語料中平均判刑 66.94 個月，恐嚇取財罪在人工標記語料中平均判刑 6.96 個月。

### 5.1 使用特徵與實驗結果

我們分別預測強盜以及恐嚇取財案件的有期徒刑（以月為單位），並取四疊交叉驗證 (4-fold cross validation)。我們以 Pearson correlation coefficient (PCC) 以及 (root-mean-square error) RMSE 做評估，並使用 Additive regression 做預測。特徵除了行為人個數外，其餘皆與案件分類時相同。

Additive regression 模型透過結合多個一維函式建立回歸表面以達到降低資料維度的目的，並且模型假設每個函式的影響皆為累加性。假設  $Y$  為預測結果， $X$  表獨立變數共  $k$  維， $f$  為可計算 lowess 的任意函式，則 Additive regression 模型為下列形式

$$Y = A + \sum_{j=1}^k f_j(B_j X_j) + \varepsilon$$

表 14. 單一特徵預測強盜量刑結果

| 特徵               | 特徵數   | PCC   | RMSE    |
|------------------|-------|-------|---------|
| Fact_tfidf       | 10034 | 0.941 | 13.9824 |
| Local label 標記次數 | 16    | 0.838 | 22.8940 |
| Local label 標記順序 | 256   | 0.913 | 16.9355 |
| 行為人特徵            | 4     | 0.474 | 36.4526 |
| 犯後態度             | 3     | 0.478 | 36.3674 |
| Behavior_tfidf   | 2897  | 0.945 | 13.5303 |
| 行為關鍵字            | 23    | 0.825 | 23.3904 |
| 被害人反應關鍵字         | 23    | 0.302 | 39.4695 |
| 被害人個數            | 3     | 0.295 | 39.5495 |
| 共犯個數             | 4     | 0.126 | 41.0622 |
| 財物               | 14    | 0.490 | 36.0889 |
| 犯案工具             | 17    | 0.607 | 32.9075 |

表 15. LOO 於預測強盜量刑結果

| 特徵               | 特徵數   | PCC   | RMSE   |
|------------------|-------|-------|--------|
| Fact_tfidf       | 3270  | 0.934 | 14.745 |
| Local label 標記次數 | 13288 | 0.957 | 11.960 |
| Local label 標記順序 | 13048 | 0.947 | 13.708 |
| 行為人特徵            | 13300 | 0.947 | 13.708 |
| 犯後態度             | 13301 | 0.947 | 13.708 |
| Behavior_tfidf   | 10407 | 0.940 | 13.708 |
| 行為關鍵字            | 13281 | 0.949 | 13.615 |
| 被害人反應關鍵字         | 13281 | 0.947 | 13.708 |
| 被害人個數            | 13301 | 0.947 | 13.708 |
| 共犯個數             | 13300 | 0.947 | 13.708 |
| 財物               | 13290 | 0.947 | 13.708 |
| 犯案工具             | 13287 | 0.947 | 13.708 |

表 16. 所有特徵及 Forward selection 於預測強盜罪量刑結果

| 選擇特徵                              | 特徵數   | PCC          | RMSE          |
|-----------------------------------|-------|--------------|---------------|
| 所有特徵                              | 13304 | 0.947        | 13.708        |
| Behavior_tfidf、Fact_tfidf、犯案工具、財物 | 12961 | <b>0.954</b> | <b>11.476</b> |

表 17. 單一特徵預測恐嚇取財罪量刑

| 特徵               | 特徵數   | PCC   | RMSE   |
|------------------|-------|-------|--------|
| Fact_tfidf       | 10034 | 0.996 | 1.688  |
| Local label 標記次數 | 16    | 0.809 | 11.060 |
| Local label 標記順序 | 256   | 0.994 | 2.105  |
| 行為人特徵            | 4     | 0.480 | 16.363 |
| 犯後態度             | 3     | 0.136 | 18.482 |
| Behavior_tfidf   | 2897  | 0.996 | 1.776  |

|          |    |       |        |
|----------|----|-------|--------|
| 行爲人關鍵字   | 23 | 0.947 | 5.998  |
| 被害人反應關鍵字 | 23 | 0.797 | 11.258 |
| 被害人個數    | 3  | 0.225 | 18.178 |
| 共犯個數     | 4  | 0.151 | 18.442 |
| 財物       | 14 | 0.627 | 14.705 |
| 犯案工具     | 17 | 0.597 | 14.968 |

表 18. LOO 預測恐嚇取財罪量刑

| 特徵               | 特徵數   | PCC   | RMSE  |
|------------------|-------|-------|-------|
| Fact_tfidf       | 3270  | 0.987 | 1.978 |
| Local label 標記次數 | 13288 | 0.996 | 1.634 |
| Local label 標記順序 | 13048 | 0.995 | 1.935 |
| 行爲人特徵            | 13300 | 0.995 | 1.935 |
| 犯後態度             | 13301 | 0.995 | 1.935 |
| Behavior_tfidf   | 10407 | 0.995 | 1.935 |
| 行爲人關鍵字           | 13281 | 0.995 | 1.935 |
| 被害人反應關鍵字         | 13281 | 0.995 | 1.935 |
| 被害人個數            | 13301 | 0.995 | 1.935 |
| 共犯個數             | 13300 | 0.995 | 1.935 |
| 財物               | 13290 | 0.995 | 1.935 |
| 犯案工具             | 13287 | 0.995 | 1.935 |

表 19. 所有特徵及 Forward selection 於預測恐嚇取財罪量刑結果

| 選擇特徵                               | 特徵數   | PCC          | RMSE         |
|------------------------------------|-------|--------------|--------------|
| 所有特徵                               | 13304 | 0.995        | 1.935        |
| Fact_tfidf、Behavior_tfidf、被害人反應關鍵字 | 12954 | <b>0.996</b> | <b>1.602</b> |

我們實驗使用單一特徵、所有特徵、Leave one out (LOO)以及以 Forward selection 找出最佳特徵組合，強盜案件之實驗結果如表 14，15 及 16，最佳組合（表 16）可達到 PCC = 0.954 及 RMSE = 11.4761 之結果。相對於平均約五年半的強盜罪刑期，我們的系

統可達到一年以下之平均預測誤差。恐嚇取財案件之實驗結果如表 17, 18 及 19, 最佳組合 (表 19) 可達到  $PCC = 0.996$  及  $RMSE = 1.6022$  之結果。相對於平均約七個月的恐嚇取財罪刑期, 我們的系統可達到兩個月以下之平均預測誤差。

## 5.2 討論

我們依據特徵值的權重, 來討論特徵值在預測量刑上的影響。特徵值的權重如表 20。為討論「非 TF-IDF 特徵」對於量刑的影響, 表 21 列出強盜及恐嚇取財案件中的「非 TF-IDF 特徵」權重。我們針對「非 TF-IDF 特徵」做討論 (如表 22), 並對於人工標記語料作統計以協助驗證。

表 20. 預測恐嚇取財與強盜量刑的前 10 項特徵值權重

| 強盜       |             | 恐嚇取財     |                |
|----------|-------------|----------|----------------|
| 權重       | 特徵          | 權重       | 特徵             |
| 30.7969  | bkey_強暴     | 56.62879 | ftfidf_易科      |
| 27.9362  | ftfidf_被害人  | 53.65348 | ftfidf_此外      |
| 15.52409 | ftfidf_保管   | 11.02504 | accomplice_3~5 |
| 13.78044 | ftfidf_當時   | 8.084648 | ftfidf_犯罪      |
| 13.12859 | ftfidf_便利商店 | 4.883647 | ftfidf_起訴書     |
| 12.0484  | ftfidf_喝令   | 3.959793 | ftfidf_正當      |
| 10.61914 | ftfidf_丙○   | 3.877146 | ftfidf_開立      |
| 7.996218 | ftfidf_不足採信 | 2.892254 | ftfidf_結果      |
| 7.991696 | ftfidf_收銀機  | 2.532131 | ftfidf_當場      |
| 7.470467 | ftfidf_得知   | 2.511225 | ftfidf_得逞      |

表 21. 預測強盜罪及恐嚇取財罪量刑非 TF-IDF 前 10 項特徵的權重

| 強盜       |             | 恐嚇取財     |                |
|----------|-------------|----------|----------------|
| 權重       | 特徵          | 權重       | 特徵             |
| 30.7969  | bkey_強暴     | 11.02504 | accomplice_3~5 |
| 0.555235 | bkey_客觀     | 0.719344 | victim_1~2     |
| 0.504492 | bkey_拿起     | 6.15E-01 | vkey_抗拒        |
| 0.503787 | property_證件 | 5.80E-01 | vkey_取財        |

|          |                 |          |               |
|----------|-----------------|----------|---------------|
| 0.439038 | bkey_心生畏懼       | 5.48E-01 | bkey_客觀       |
| 0.406706 | tool_棒棍         | 5.32E-01 | bkey_強行       |
| 0.38426  | property_手機電子用品 | 0.509546 | tool_槍        |
| 0.351021 | property_五千到一萬  | 0.328142 | property_一千以下 |
| 0.325931 | tool_槍          | 0.110464 | property_其他   |
| 0.322338 | property_皮包     | 5.92E-02 | bkey_恐嚇       |

表 22. 量刑預測特徵討論

| 特徵       | 強盜   | 恐嚇取財                                       |
|----------|--|--|
| 犯後態度     | 若未坦承犯行平均判刑 100.5 個月  | 若未坦承犯行平均判刑 7.26 個月                         |
| 行為人特徵    | 有前科平均判刑 97.1 個月<br>有精神疾病平均判刑 53.3 個月                         | 有前科平均判刑 10.35 個月<br>累犯平均判刑 7.45 個月         |
| 行為關鍵字    | 包含強暴、客觀、拿起、心生畏懼、交付、抗拒、不能、取走、脅迫、意圖等關鍵字平均判刑 81.08 個月           | 包含客觀、強行、恐嚇、身體、拿起等關鍵字平均刑期為 11.5 個月          |
| 被害人反應關鍵字 | 包含抗拒、拿起、心生畏懼等關鍵字平均判刑 71.62 個月                                | 包含抗拒、取財、心生畏懼、交付等關鍵字平均判刑 10.91 個月           |
| 犯案工具     | 使用棒棍、槍、交通工具、刀械、縱火工具等工具平均判刑 72.07 個月                          | 使用槍、交通工具、刀械、信函、言語行為、道具槍、其他等工具平均判刑 15.52 個月 |
| 財物       | 財物包含證件、手機電子用品、五千到一萬、皮包、一千以下、珠寶金飾、一千到五千、百萬以上、其他等平均判刑 72.23 個月 | 財物包含一千以下、十萬到百萬、一萬到十萬等平均判刑 9.95 個月          |
| 共犯個數     | 共犯 5 以上、1~3 人平均判刑 49 個月                                      | 共犯 3~5 人、1~3 人平均判刑 8.46 個月                 |
| 被害人個數    | 被害人數 1~2、3~5 人平均判刑 74 個月                                     | 被害人數 1~2 人平均判刑 8.53 個月                     |

## 6. 以自動標記進行案件分類及量刑預測

由於人工標記法律要素標籤需大量人力與時間，因此我們以第三節中介紹之自動標記系統，取代第四節及第五節中人工標記，以分別進行案件分類及量刑預測。本研究收集之 2113 件案件中，140 件已進行人工標記（請參考第三節），其餘 1973 件由第三節所述的自動標記系統進行標記。由於 Local label 標記結果不甚理想，因此我們僅採用「法官認定事實」以及「行為」資訊進行案件分類及量刑預測，取四疊交叉驗證 (4-fold cross validation) 分別以 F1 score、PCC 以及 RMSE 做評估，表 23、24 分別列出自動標記在分類及量刑預測的實驗結果。

表 23. 自動標記案件分類結果

| 特徵             | 特徵數   | F1 score     |
|----------------|-------|--------------|
| Fact_tfidf     | 10034 | 0.797        |
| Behavior_tfidf | 2897  | 0.592        |
| 行為關鍵字          | 23    | 0.513        |
| 所有特徵           | 12954 | <b>0.801</b> |

表 24. 自動標記案件量刑預測結果

| 特徵             | 特徵數   | 強盜           |                | 恐嚇取財         |               |
|----------------|-------|--------------|----------------|--------------|---------------|
|                |       | PCC          | RMSE           | PCC          | RMSE          |
| Fact_tfidf     | 10034 | 0.073        | 114.251        | 0.019        | 45.812        |
| Behavior_tfidf | 2897  | 0.067        | 117.284        | 0.016        | 49.533        |
| 行為關鍵字          | 23    | -0.015       | 127.017        | -0.003       | 51.074        |
| 所有特徵           | 12954 | <b>0.085</b> | <b>112.144</b> | <b>0.022</b> | <b>44.823</b> |

## 7. 結論

本研究針對強盜罪與恐嚇取財罪提出一自動標記法律要素標籤，並進一步分類案件與預測量刑之系統，以期能提供法官做為判決參考。首先，我們定義強盜罪與恐嚇取財罪之法律要素標籤來闡述法律文件中的犯罪資訊，並試圖以分類器自動標記標籤。其次，透過法律要素標籤資訊，我們可以改善案件分類以及量刑預測之效果。最後，我們探討影響案件分類以及量刑預測的法律資訊，並加以統計佐證。關於本研究未來可能之延伸方向有三：(1) 我們認為自動標記 local label 的部分，仍有極大的改善空間。(2) 考慮重疊標記以及加入法官資訊，將會有助提升案件分類及量刑預測之準確性。(3) 關於法律要素標籤資訊用於案件分類與量刑預測方面，也可佐以司法院之統計資料，以作更詳細之比較。

## 誌謝

特別感謝林中鶴、孫斌、潘佑達、林執中等先進為本研究提供極大的協助。

## 參考文獻

- Andreas, H., Staab, S. & Stumme, G. (2003). Ontologies Improve Text Document Clustering. *IEEE International Conference on Data Mining*.
- Ashley, K. D. & Bruninghaus, S. (2009). Automatically Classifying Case Texts and Predicting Outcomes. *AI and Law*, 17.
- Jiang, W., Wang, X.-L. & Guan, Y. (2006). Improving Sequence Tagging using Machine Learning Techniques. *International Conference on Machine Learning and Cybernetics*.
- Kuo, H.-C., Tsai, T.-H. & Huang, J.-P. (2006). Building a Concept Hierarchy by Hierarchical Clustering with Join/Merge Decision. *Joint Conference on Information Sciences*.
- Lame, G. (2001). A Categorization Method for French Legal Documents on the Web. *International Conference on Artificial Intelligence and Law*, 219-20.
- Liu, C.-L., Chang, C.-T. & Ho, J.-J. (2004). Case Instance Generation and Refinement for Case-Based Criminal Summary Judgments in Chinese. *Journal of Information Science and Engineering*, 20, 783-800.
- Schild, U. (1998). Criminal Sentencing and Intelligent Decision Support. *AI and Law*, 6, 151-202.
- Wibowo, W. & Williams, H. E. (2002). Simple and Accurate Feature Selection for Hierarchical Categorisation. *ACM Symposium on Document Engineering*.
- Zhang, C., Xu, X. & Zhang, C. (2008). Analysis of the Factors Affecting the Performance of CRF-based Keywords Extraction Model. *New Technology of Library and Information Service*, 24, 34-40.

