

應用分類樹於兩車肇事事故鑑定專家支援系統之建構

葉名山¹ 劉欣憲² 張恭文³ 陳鵬升⁴

摘要

本研究採用分類樹之技術構建兩車肇事事故專家支援系統，欲協助鑑定人員有一快速且可供參考之流程，提升事故鑑定速度與解決經驗傳承不良之問題。本研究採用資料庫為 86 年至 93 年期間覆議會與車鑑會鑑定結果一致之案件，資料共 1295 件(2590 筆當事者資料)，然再透過所有事故類型混合、汽車對汽車(588 筆案件數；1176 筆當事者數)、以及非汽車對汽車(含汽車碰撞機車，資料共 707 筆案件數；1414 筆當事者數)等三種事故類型，輔以雙方變數與單方變數構建出 6 個主要模式，並以決策樹之分支節點所要求樣本數不同，加以細分成 24 個子模式。本研究結果顯示，以所有事故類型混合輔以雙方變數進行構建之分類樹模式，且分支節點最小要求的數目為 50 個樣本數時為最佳模式，該模式訓練的判中率有 64.77%，而驗證之判中率有 71.43%。

壹、前言

近年來，臺閩地區肇事案件呈現增長之趨勢，在 93 年仍有 134,194 的道路交通事故案件數，再與歷年相比依舊無減緩之狀況，輔以近年國人日漸重視自身相關權益，因道路交通事故發生後，肇事責任判定對於兩造後續賠償、保險與民事理賠息息相關，故導致申請肇事鑑定案件居高不下。於 93 年申請鑑定之案件數達到 18,250 件，與 92 年相比仍高出 9.2%；然以我國目前台灣地區有十四個各地區車輛行車事故鑑定委員會(車鑑會)與三個台灣省車輛行車事故覆議委員會(覆議會)，審議全國每年近兩萬件申請之交通事故鑑定案，其負擔相當沉重，易造成鑑定時程延宕過久與效率不彰之議；此外，國內長年鑑定人員人事更迭與鑑定準則不一，亦會造成肇事鑑定工作無良好傳承制度以及類似案件類型所鑑定結果不同，遂有鑑定品質不良之議。因此在我國車輛行車事故鑑定制度中，有著交通事故筆錄資料品質不佳、事故鑑定人員之專業能力不足之問題。為解決經驗傳承不佳與引用鑑定準則不同之狀況，以及上述各種鑑定之負面影響，因此制定一套標準鑑定流程，供後續鑑定委員有一套鑑定基準可循，希可減少鑑定時程過久與經驗傳承不良之狀況。因探討交通肇事文獻多以統計方法為主，其中以迴歸分析進行預測為最常見者，然在傳統線性迴歸、卜瓦松迴歸、及負二項迴歸之預測能力亦不盡理想，故本研究採用資料探勘之分類樹方法進行構建兩車肇事事故鑑定之專家支援系統。本研究旨在採用分類樹方法構建一套道路交通事故責任判定系統，提供車輛肇事事故鑑定人員於進行肇責判定時有較可靠之支援系統可供參考，並使鑑定流程簡單化與快速化以加強該系統之參考依據。

1 逢甲大學交通工程與管理學系副教授。
2 逢甲大學交通工程與管理學系碩士班研究生。
3 逢甲大學交通工程與管理學系大學生。
4 逢甲大學交通工程與管理學系碩士班研究生。

貳、文獻回顧

2.1 分類樹之簡介

在資料探勘之技術中，以預測技術著名的乃是決策樹(decision tree)與類神經網路(artificial neural network)為主，但因類神經網路在進行預測過程中有黑箱作業之議，判定過程無法明白呈現，故要把紀錄分群或預測結果時，欲將整體過程輸出且易於判別而言，決策樹法為一個較佳之選擇。一般而言決策樹不一定比其他模式之預測能力精準，但是與其他模式相比，決策樹較易使人了解判別過程[1]。而在決策中有兩種不同型態的樹狀，分別為分類樹(classification tree)與迴歸分析樹(regression tree)，分類樹是在將標記紀錄分派到適當之類別中，並提供正確分類之可靠度；而迴歸分析樹乃是在估計某一變數之目標值[2]。本研究將各類肇事型態歸納至五種判定責任(全部原因、主要原因、同為原因、次要原因、以及無肇事原因)，故本研究採用分類樹法進行肇事鑑定之模式構建。

分類樹進行分類之演算法與分類技術陸續被提出，然常見有 ID3(與之後發展的 C4.5)、Gini Index、Twoing 以及 PRISM，其演算法之過程要述於下：

- 一、ID3：此演算法是由 Quinlan 所提出，ID3 在決策樹歸納過程中使用了 Shannon 於 1949 年所提出的資訊理論之中的資訊獲利(Information Gain)做為屬性之測試依據。資訊獲利是當某一屬性為決策樹節點時，其分支節點的加權熵值(Entropy)與原物件集合熵值的差距。應用於決策樹中之某一物件集合 C 分屬於 j 個不同類別，則此物件集合的熵值 $E(C)$ 公式為：

$$E(C) = - \sum_j P_j \times \log_2 P_j \dots\dots\dots (式 1)$$

其中 $P_j = (\text{屬於類別 } j \text{ 的物件數}) / (C \text{ 的總物件數})$

再以屬性 A_i 為決策樹節點時，該節點以下建立了 m 個子節點，其分支節點的加權熵值計算如下式 2：

$$E(A_i) = \sum_{k=1}^m (n_k / n) \times E(C_k) \dots\dots\dots (式 2)$$

其中 C_k 為物件集合 C 中其屬性值相同的物件子集合 k ；

$E(C_k)$ 為物件子集合 C_k 的熵值；

n 為物件集合 C 的總物件數；

n_k 為物件子集合 C_k 的物件數。

茲因資訊獲利是指選擇屬性 A_i 為決策樹節點時帶來熵值的改變程度，也就是原來物件集合的熵值，與以 A_i 為決策樹節點時，其分支節點加權熵值的差距，即 $G(A_i) = E(C) - E(A_i)$ ，最後在選取資訊獲利最大者作為選取之屬性。

- 二、C4.5：該 C4.5 演算法是由 ID3 演算法改良獲得，因 ID3 以資訊獲利為選擇分類屬性的準則，偏向選擇屬性值較多的的屬性。遂 C4.5 改以獲利比 (Gain-Ratio Measure) 作為選擇屬性的衡量準則，如式 3 所示：

$$GR(A_i) = G(A_i) / IV(A_i) \dots\dots\dots (式 3)$$

其中 $GR(A_i)$ 屬性 A_i 對 C 物件集合進行分割的資訊獲利；

$$IV(A_i) \text{ 為 } A_i \text{ 屬性的資訊值} = - \sum_m P_m \times \log_2 P_m ;$$

p_m 為 C 物件集合中，屬性之值為 a_m 的物件數 / C 的物件數。

三、 Gini Index：在 CART 的分類法中，利用 Gini Index 之目的在於評估選取何種屬性能將節點的多樣性(Diversity)降低最多。然根節點的 Gini Index 與其爾後在進行分割之後加權的 Gini Index 應如式 4 與式 5。再將 $Gini(t)$ 扣除 $Gini'(t)$ 之值，但是該屬性為分隔條件所降低之分散度，該選取較大者為佳。然式 4 與式 5 如下所示：

$$Gini(t) = 1 - \sum_i p_i^2 \dots\dots\dots(式 4)$$

其中 t 為已知節點 t ；

P_i 為第 t 個節點中第 i 個類別之機率。

$$Gini'(t) = \frac{N_1}{N} Gini(t_1) + \frac{N_2}{N} Gini(t_2) \dots\dots\dots(式 5)$$

其中支點 t 切割成 t_1 和 t_2 ；

N_1 、 N_2 為分割成 t_1 、 t_2 中的樣本。

四、 Twoing：該演算法其原先的概念與 Gini Index 不同，是在於計算父節點與子節點間最大的不純度改善值，目的是將所有的類別區分成兩大類。當一個目標變數(Target variable)有類別 $C = \{1, 2, 3, \dots, J\}$ ，將所有類別分成 2 個大類別如： $C_1 = \{j_1, j_2, \dots, j_n\}$ ，則 $C_2 = C - C_1$ 。然在 Twoing 進行分割時，將類別 C 分成 C_1 與 C_2 之動作視為一個二類別問題，因此於節點 t 之不純度改善值 $\Delta i(s, t)$ 的大小，茲受到類別 C_1 影響，遂將不純度改善值改寫為 $\Delta i(s, t, C_1)$ 。然而 Twoing 定義的不純度函數如式 6 所示：

$$\Phi(s, t) = \frac{P_L P_R}{4} \left[\sum_j |p(j|t_L) - p(j|t_R)| \right]^2 \dots\dots\dots(式 6)$$

因 Twoing 在分支時，並非評估單一節點之不純度，而以父節點與子節點之間所能改進的最大不純度值，而該分類方式與 ID3 演算法的 Entropy 值較為類似。

五、 PRISM：該演算法於 1987 年提出，其基本概念與 ID3 相似，但 PRISM 以屬性值配對做為分類的依據；不過仍依資訊獲利為主要概念，而資訊獲利之公式改以條件機率做計算，公式如下所示：

$$PRISM_Gain(A=x) = \log_2 \left(\frac{p(A=x|P)}{p(A=x)} \right), \text{ 當 } P(A=x|P) \neq 0 \dots\dots\dots(式 7)$$

其中 $p(A=x)$ 代表所有樣本中符合 $A=x$ 屬性值配對的樣本個數；

$p(A=x|P)$ 代表正例中符合 $A=x$ 屬性值配對的樣本個數。

決策樹演算方法與軟體技術發展已應用相當廣泛，但決策樹在國內交通肇事研究中較少被採用，遂本研究針對決策樹應用於交通方面與非交通領域兩部份進行文獻評析，目前有運用決策樹演算法及多元羅吉特模式，分析影響交通事故嚴重程度之主要原因與影響程度，從中得知造成交通事故嚴重程度之因素，並加以評估改良之研究[3]；尚有應用於決策樹演算法及負二項迴歸模式，分析高速公路肇事次數形式之資料，以得知事故發生之關鍵因素，預防改善之研究[4]。但在近幾年來其他研究領域中，決策樹被廣泛運用在各研究領域上，包含電信、醫療、企業管理與地理資訊系統等部分。電信部份有應用決策樹歸納法，探討台灣具有影響高價值行動電話用戶的相關屬性與特徵之因素，作為未來市場區隔的基礎及區隔後市場的描繪，以提供行動電話系統業者制定行銷策略作為參考依據[5]；然在醫療部份，有應用決策樹之 C4.5 演算法和羅吉斯迴歸建立模式，分析比較精神

分裂症之病人與正常人在認知功能缺損的情形[6]，作為未來防治之要項；在企業管理方面，有應用決策樹中的 C4.5 演算法，分析並探討建設公司之售後服務資料庫，並對於研究結果之工程品質不良因素加以改良[7]；而在地理資訊系統運用模糊理論及決策樹法於空載影像分類，探討自行設定之分類規則運用於不同地區的歷史空載影像圖的實用性，並利用各分類準則於空載影像進行分類之適用性[8]。

2.2 影響肇事鑑定之變數

在交通肇事文獻中，多以迴歸分析與程序性判別方法進行研究，然在迴歸分析挑選變數方面，先以主觀認定該變數採用之合適性在進行顯著分析，以求影響肇責之顯著變數，遂迴歸分析中所挑選出之變數，亦可供本研究之參考依據；在程序性驗證分析方面，所考量流程之相關因素以判定肇事責任，亦可得到明顯影響肇責之因素，故本研究可採納諸多研究中之選取變數。由上述可知，挑選變數進行肇事分析亦是重要一環，遂彙整諸多研究所選取之變數於表 1，供本研究於後續分析之中作為參考之用。

表 1、肇事分析相關文獻彙整表

作者(年代)	研究目的	研究方法	選取變數
邱裕鈞、郭曜彰 民 94[9]	建置汽機車碰撞事故之鑑定決策支援系統	類神經網路	超速、飲酒、天候、道路類型、道路型態、速限、動作、行向、車損部位、筆錄速率、道路優先權
邱裕鈞、方守潔 民 93[10]	建置兩車碰撞事故之鑑定決策支援系統	判別分析、類神經網路	區別、車種、超速、飲酒、道路類別、道路型態、行向、車損部位、筆錄速率、是否預見、駕亡、路權
邱裕鈞諸君 民 93[11]	建置兩車碰撞事故之鑑定決策支援系統	階層式群落分析、類神經網路	車種、超速、飲酒、道路類別、道路型態、行向、車損部位、筆錄速率、是否預見、駕亡、路權
葉名山、林世淵 民 93[12]	建置事故鑑定型與理賠機制	專家問卷調查與權重分析	行向、左右方區別、超速、飲酒
楊宗璟諸君 民 93[13]	研究直行與同向右轉車發生碰撞致使一方無肇事責任之模式	羅吉特模式	動作、碰撞位置、預見、煞車、超速、飲酒、採取措施、車種、筆錄速率、傷亡
楊宗璟、林芳誼 民 93[14]	構建路段中變換車道者與同向直行車發生碰撞事故責任與傷亡統計模式	多元羅吉特模式	駕駛動作(路邊起駛、變換車道或超車型態)、超速狀況、飲酒程度、車損部位、是否預見、預見距離
楊宗璟、陳婉郁 民 92[15]	構建路口橫向碰撞事故鑑定分析之模式	羅吉特模式	右方車是否靠右行駛、車損部位(雙方變數)、是否超速、是否飲酒、是否預見、駕駛者採取措施
陳高村、廖信智 民 91[16]	對於路段肇事事故構建鑑定準則	因果分析與系統化樹狀圖	碰撞型態、事故時間、事故地點、行車方向

表 1、肇事分析相關文獻彙整表(續)

楊宗璟、艾嘉銘 民 91[17]	分析交通違規事故責任鑑定影響因素	統計分析方法	事故類型、車種、是否超速與其程度、是否飲酒與其程度、天色、天候、當事者雙方之相對位置、兩車動作、碰撞方向
張漢威 民 90[18]	解釋行車肇事於鑑定作業時對於肇事責任歸屬所考量之因素	因素分析圖	行向、碰撞類型、道路型態、駕駛行為、相關違規事項
張漢威 民 90[19]	構建肇責因素分析表	系統化整理	路況、駕駛行為、違規行為、禮讓之行為
周文生、陳蔚文 民 88[20]	以交通法規引用為導向構建肇事鑑定之原因分析的原則與方法	決策樹	行駛方向、路權優先順序、碰撞型態、道路型態及路況、駕駛行為、綜合分析肇事因果關係
吳宗修 民 87[21]	以路權為基準構建非號誌化交岔路口事故責任鑑定原則	模擬分析與圖解法	路權優先權、車輛行向、車流遭遇型態
陳高村 民 87[22]	建置不同碰撞型態行為之事故原因分析與肇事責任鑑定法則	系統化整理與圖解法	碰撞型態(對撞、角撞、側撞與追撞)、行車方向、碰撞軌跡、事故發生位置
陳高村、龍天立 民 84[23]	侵權形態事故肇事責任鑑定方法之構建	法理適用原則分析	車輛尺度、碰撞位置、車損擦痕位置、事故型態、路權

參、資料說明統計與變數選取

本章針對本研究所採用之肇事資料庫進行變數說明與進行變數選取加以討論，且在最後將選取出之顯著變數進行資料基本統計，其說明如后。

3.1 資料說明

本研究所採用之資料庫為 86 年至 93 年期間，台灣地區道路交通事故案件已建檔入資料庫之資料中，萃取可用肇事事案件數共 1295 件(共 2590 筆當事者資料)，該資料庫之建檔變數共計分成共同狀況、當事者資料、以及當事者行動狀態等三大部分，所建檔之相關變數屬性說明茲列於后：

共同狀況：1.案號、2.肇事時間(年、月、日、時、分)、3.肇事時段、4.肇事縣市、5.事故位置型態、6.行向關係、7.天候、8.光線、9.速限、10.號誌種類、11.號誌動作、12.道路類別，在第一類型中共有 12 個變數。在案號、肇事時間(年、月、日、時、分)、肇事縣市等部分，此三項變數於交通事故鑑定之狀況屬於共同情形，且不足以影響肇事責任判定，故本研究不採納上述變數；而行向關係及速限兩者，分別與兩車關係及超速資料重疊性過大，因此本研究亦不採納此兩項變數。故在共同狀況初步篩選出 7 項變數。

當事者資料：1.當事人出生時間(年、月、日)、2.年齡、3.車種、4.性別、5.受傷程度、6.保護裝置、7.駕駛資格情形、8.駕駛人飲酒情況、9.肇事因素、10.主要肇事因素、11.車輛碰撞部位，在第二類型中共有 11 項變數。因本研究已將當事人

出生時間(年、月、日)轉換成年齡；而性別變數本研究認定與肇責判定無直接關聯；然受傷程度屬於結果變數；再者因肇事因素、主要肇事因素於警員填寫資料時之不全，所以資料紊亂且認定不一；因此本研究不考慮上述 5 項變數。故在當事人資料初步篩選出 6 項變數。

當事者行動狀態：1.超速、2.筆錄速度、3.行車方向、4.動作、5.原所在車道數、6.原所在車道位置、7.轉後車道數、8.轉後車道位置、9.兩車關係、10.是否過路口中心、11.違規行為、12.幹道或支道、13.閃光號誌、14.當事人行動、15.肇事逃逸、16.事故類型及形態、17.變換車道，在第三類型中共有 17 項變數。茲因筆錄速度已轉換成超速變數使用，且行車方向與兩車關係重疊性較大；再者原所在車道數、原所在車道位置、轉向後車道數、轉後車道位置等變數，本研究因其變數相當複雜，依據現有資料甚難研判其對肇事責任之因果關係；再者因類別變數過多的變數有當事人行動(計有 21 類)、事故類型及形態(計有 36 類)，將使分類樹在分類進行時，預期產生過度訓練之狀況；因此本研究將上述 8 項變數皆不考量使用。故在當事者行動狀態初步篩選出 9 項變數。

3.2 變數選取

本研究採用交叉分析之卡方檢定進行重要變數屬性選取。交叉分析表可以顯示兩個以上聯合頻率分配表，因此又稱之為多維度頻率分配表。然交叉分析表對資料處理除了有分類資料功能外，尚可進行簡單的統計推定，故本研究針對變數篩選部分，以軟體 SPSS 10.0 之交叉分析表功能進行卡方檢定進行挑選顯著變數。

本研究進行交叉分析表所計算之 P 值小於 0.05 為顯著變數之挑選依據，然其所挑選之顯著變數有 1.年齡、2.車輛碰撞部位、3.閃光號誌、4.是否過路口中心、5.保護裝置、6.事故位置型態、7.號誌動作、8.道路類別、9.號誌種類、10.車種、11.兩車關係、12.違規行駛、13.幹道或支道、14.超速、15.駕駛人飲酒情況、16.動作、以及 17.變換車道等十七項變數。本研究由第一階段所篩選出之二十二項變數中肇事時段($P=0.066$)、天候($P=0.101$)、光線($P=0.252$)、駕駛資格情形($P=0.055$)與肇事逃逸($P=0.133$)共五項變數為不顯著變數，故本研究採用十七項變數之數據資料進行決策樹模式之訓練與驗證用。

3.3 選取變數資料統計

本研究所採納之變數經交叉分析篩選後，所得之變數與該分類變項之基本統計量與所佔百分比，茲整理於下表所示：

表 2、變數名稱與分類之內容彙整表

序號	名稱	類別數	內容
1	肇事責任	5	1.完全原因(644 筆, 佔 24.9%)、2.主因(573 筆, 佔 22.1%)、3.同為原因(156 筆, 佔 6.0%)、4.次因(573 筆, 佔 22.1%)、5.無因(644 筆, 佔 24.9%)
2	年齡	5	1.20 歲以下(139 筆, 佔 5.4%)、2.20~35 歲(1115 筆, 佔 43.1%)、3.35~50 歲(902 筆, 佔 34.8%)、4.50~65 歲(313 筆, 佔 12.1%)、5.65 歲以上(121 筆, 佔 4.7%)
3	車輛碰撞部位	5	1.前車頭(811 筆, 佔 31.3%)、2.後車尾(126 筆, 佔 4.9%)、3.右側車身(677 筆, 佔 26.1%)、4.左側車身(746 筆, 佔 28.9%)、5.不明(227 筆, 佔 8.8%)
4	閃光號誌	5	1.閃紅(110 筆, 佔 4.2%)、2.閃黃(175 筆, 佔 6.8%)、3.無閃光號誌(1931 筆, 佔 74.6%)、4.行車管制號誌(336 筆, 佔 13.0%)、5.不明(38 筆, 佔 1.5%)
5	車輛是否過路口中心	3	1.否(562 筆, 佔 21.7%)、2.是(1015 筆, 佔 39.2%)、3.非路口(1013 筆, 佔 39.1%)
6	保護裝置	3	1.戴安全帽或繫安全帶(1077 筆, 佔 41.6%)、2.未戴安全帽或繫安全帶(1123 筆, 佔 43.4%)、3.不明或其他(390 筆, 佔 15.1%)
7	事故位置型態	5	1.路段(981 筆, 佔 37.9%)、2.行車管制號誌路口(446 筆, 佔 17.2%)、3.閃光號誌路口(336 筆, 佔 13.0%)、4.無號誌路口(813 筆, 佔 31.4%)、5.其他(14 筆, 佔 0.5%)
8	號誌動作	3	1.正常(850 筆, 佔 32.8%)、2.非正常運作(124 筆, 佔 4.8%)、3.無號誌(1616 筆, 佔 62.4%)
9	道路類別	6	1.國道(56 筆, 佔 2.2%)、2.省道(584 筆, 佔 22.5%)、3.縣道(414 筆, 佔 16.0%)、4.鄉道(308 筆, 佔 11.9%)、5.市區道路(859 筆, 佔 33.2%)、6.其他道路(369 筆, 佔 14.2%)
10	號誌種類	4	1.行車管制號誌(460 筆, 佔 17.8%)、2.閃光號誌(394 筆, 佔 15.2%)、3.無號誌(1634 筆, 佔 63.1%)、4.不明(102 筆, 佔 3.9%)
11	車種	4	1.小車 (1577 筆, 佔 60.9%)、2.大車(176 筆, 佔 6.8%)、3.機車 (799 筆, 佔 30.8%)、4.其他(38 筆, 佔 1.5%)
12	兩車關係	8	1.同向前車(275 筆, 佔 10.6%)、2.同向後車(288 筆, 佔 11.1%)、3.同向左方車(166 筆, 佔 6.4%)、4.同向右方車(171 筆, 佔 6.6%)、5.橫向左方車(441 筆, 佔 17.0%)、6.橫向右方車(427 筆, 佔 16.5%)、7.對向(796 筆, 佔 30.7%)、8.路邊起駛(26 筆, 佔 1.0%)
13	幹道或支道	4	1.同為幹支道(2032 筆, 佔 78.5%)、2.幹道(195 筆, 佔 7.5%)、3.支道(192 筆, 佔 7.4%)、4.不明(171 筆, 佔 6.6%)

表 2、變數名稱與分類之內容彙整表(續)

序號	名稱	類別數	內容
14	違規行駛	6	1.入侵對向車道(126 筆, 佔 4.9%)、2.逆向(29 筆, 佔 1.1%)、3.無違規(1432 筆, 佔 55.3%)、4.不明(558 筆, 佔 21.5%)、5.未行使於指定車道上(121 筆, 佔 4.7%)、6.未依標誌標線行駛(324 筆, 佔 12.5%)
15	超速	4	1.嚴重超速(73 筆, 佔 2.8%)、2.超速(227 筆, 佔 8.8%)、3.無(1836 筆, 佔 70.9%)、4.不明(454 筆, 佔 17.5%)
16	駕駛人飲酒情況	3	1.合格(2337 筆, 佔 90.2%)、2.不合格(132 筆, 佔 2.1%)、3.其他或不明(121 筆, 佔 4.7%)
17	動作	7	1.直行(2054 筆, 佔 79.3%)、2.右轉(99 筆, 佔 3.8%)、3.左轉(275 筆, 佔 10.6%)、4.迴轉(76 筆, 佔 2.9%)、5.靜止(71 筆, 佔 2.7%)、6.倒車(14 筆, 佔 0.5%)、7.不明(1 筆, 佔 0.02%)
18	變換車道	4	1.無(2198 筆, 佔 84.9%)、2.有(309 筆, 佔 11.9%)、3.超車(49 筆, 佔 1.9%)、4.不明(34 筆, 佔 1.3%)

肆、模式構建

本研究利用交叉分析表之卡方檢定後篩選出顯著變數, 以供 IBM DB2 Intelligent Miner 軟體建構決策樹變數之用。然再將事故類型分成三類, 依序有所有事故類型混合、汽車對汽車、與非汽車對汽車等三大大事故類型, 並配合採用資料之形式是以雙方變數(案件數)或者是(當事者)為主, 構建出六個決策樹模式。在輸入變數方面, 茲因汽車對汽車與非汽車對汽車事故類型比所有事故類型混合少一項變數為「車種」, 故於構建決策樹模式之輸入變數上會有差異; 而輸出變數方面, 雙方變數之輸出原應有 2 個輸出變數, 意即應有當事者一與當事者二之肇事責任, 但因考量軟體應用之限制與預測單一當事者之肇責及形同預測兩個當事者之肇責, 故於案件數上僅考量當事者一之肇事責任。本研究將軟體構建出最大樹(Maximal Tree)後採用盆栽法進行裁減加以修剪(Pruning), 利用分支節點樣本數限制為 50、100、150、以及 DB2 Intelligent Miner 軟體自行修剪之最適樹, 構建出以六個模式為主輔以四個修剪之後之限制, 構建出一個較為完善的模式架構, 如表 3 所示。然以六個模式輔以四個限制構建出共 24 個子模式, 然該子模式群將於後續驗證中, 挑選訓練與驗證結果最佳者, 採用為本研究之肇事鑑定系統決策樹模式。

而模式構建與驗證的資料筆數方面, 所有事故類型共有 1295 件(共 2590 筆當事者資料)、汽車對汽車碰撞之 588 筆案件數(共 1176 筆當事者數)與非汽車碰撞之 707 筆案件數(共 1414 筆當事者數), 各模式資料筆數所佔之 80%進行訓練與 20%進行驗證, 以便求得最佳的樹狀模型, 驗證結果分析於下一章節。

表 3、分類樹各模式之相關參數選取彙整表

相關變數說明					分類樹分支節點樣本數限制			
變數挑選方式	事故類型	訓練資料依據	輸入變數個數	輸出變數個數	50 個	100 個	150 個	修剪最適樹
交叉分析表所選顯著變數	所有事故類型混合	雙方變數	34	1	模式 1-a	模式 1-b	模式 1-c	模式 1-d
		單方變數	17	1	模式 2-a	模式 2-b	模式 2-c	模式 2-d
	汽車對汽車事故類型	雙方變數	32	1	模式 3-a	模式 3-b	模式 3-c	模式 3-d
		單方變數	16	1	模式 4-a	模式 4-b	模式 4-c	模式 4-d
	非汽車對汽車之事故類型	雙方變數	32	1	模式 5-a	模式 5-b	模式 5-c	模式 5-d
		單方變數	16	1	模式 6-a	模式 6-b	模式 6-c	模式 6-d

註：雙方變數之輸入變數個數=單方變數之輸入變數個數×2；

雙方變數之輸出變數個數為當事者一之肇事責任。

伍、模式驗證

5.1 判中率分析

本研究針對所構建出各子模式之決策樹進行模式驗證之動作，然模式驗證出結果如下表 4 所示：

表 4、分類樹各子模式驗證分析表

類型	節點限制數	雙方變數				單方變數			
		模式	訓練判中率	驗證判中率	模式順序	模式	訓練判中率	驗證判中率	模式順序
所有事故類型混合	50	1-a	64.77%	71.43%	1	2-a	59.22%	55.41%	3
	100	1-b	57.91%	66.02%	2	2-b	57.38%	56.95%	2
	150	1-c	54.92%	58.69%	4	2-c	55.50%	58.69%	1
	最適樹	1-d	60.71%	61.00%	3	2-d	54.54%	53.67%	4
汽車與汽車事故	50	3-a	64.97%	58.12%	1	4-a	61.74%	56.17%	3
	100	3-b	54.35%	54.70%	3	4-b	59.51%	62.55%	1
	150	3-c	54.35%	54.70%	3	4-c	57.07%	58.72%	2
	最適樹	3-d	58.60%	58.12%	2	4-d	53.88%	54.89%	4
非汽車事故	50	5-a	62.37%	60.28%	1	6-a	67.11%	57.60%	1
	100	5-b	54.59%	54.61%	2	6-b	55.88%	49.47%	4
	150	5-c	52.12%	51.77%	3	6-c	50.84%	56.19%	2
	最適樹	5-d	58.95%	48.96%	4	6-d	52.87%	51.24%	3

由上表 4 可知，本研究成果於所有事故類型混合方面，依相同節點限制數所構建出之分類樹模式中，以雙方變數之模式皆優於單方變數模式，故所有事故類型混合之雙方變數模式進行預測肇責較為準確；而該模式於輸入變數方面，雙方

變數以判中率 71.43% 優於單方變數判中率 58.69%。汽車與汽車事故類型方面，該類型於輸入變數方面，單方變數以判中率 62.55% 優於雙方變數判中率 58.12%。非汽車與汽車事故類型方面，該類型於輸入變數方面，雙方變數以判中率 60.28% 優於單方變數判中率 57.60%。於雙方變數所構建模式方面，三種事故類型皆有將分類樹越細分，則判中率越精準之趨勢。

本研究於最佳判中率方面，因修剪樹的分支節點進行要求最小樣本數，故訓練出模式的判中率之代表性略顯不足，但考量簡化流程與不使其過度訓練之狀況下，則以所有事故類型混合輔以雙方變數構建之模式，配合分支節點限制樣本為 50 時，訓練判中率達 64.77%，有 71.43% 的最佳驗證判中率，優於汽車與汽車事故類型之判中率 62.55%，以及非汽車事故類型之判中率 60.28%，顯示已經有相當之預測能力，因此該模式為本研究所選擇之最佳模式，其相關圖表詳述如圖 1。

5.2 最佳模式

本研究所採用分類樹之最佳模式樹狀圖，如圖 1 所示。本研究所採案件雙方變數進行預測案件肇責歸屬，故須先將案件相關變數內容至表 2 轉換成所屬代號，該模式判別方式採由根節點至末端節點(由上而下)之方式進行分類預測，圖 1 中以圓形方式代表該節點為分支節點，每個分支節點代號將對應至表 5 之屬性進行屬性值之分割動作，倘若欲進行預測之案件與進行分支之節點屬性值相同者，則分類方向向左，反之向右。當分支進行至以矩形方式表示之末端節點時，則將矩形中所屬代號對應至表 6 求得最後末端節點所代表標記，該標記則為案件所預測出之結果，以及該節點標記所代表之純度。

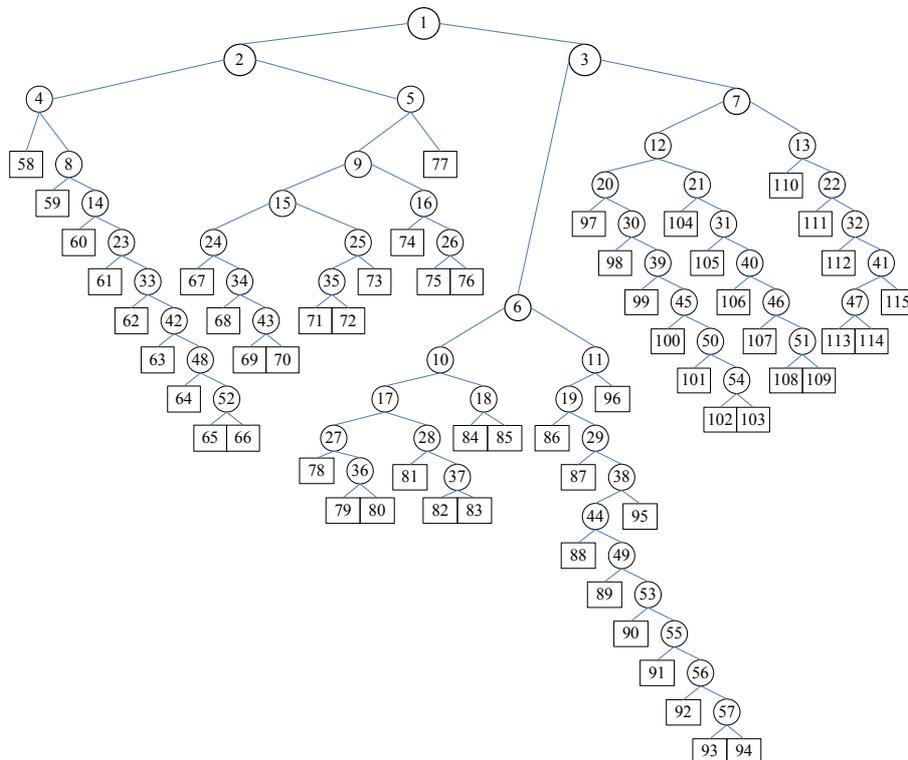


圖 1、最佳模式之分類樹狀圖

表 5、最佳模式之分支節點屬性分割與代號彙整表

節點	1	2	3	4	5	6	7	8	9
屬性	兩車關係 1	幹道或支 道 2	違規行駛 1	車種 1	動作 1	違規行駛 2	違規行駛 2	號誌種類 1	幹道或支 道 2
屬性值	5,6	2	3	2,4	1,7	3,5	2,3	1	1,4
節點	10	11	12	13	14	15	16	17	18
屬性	動作 1	兩車關係 2	違規行駛 1	動作 1	年齡 1	車輛碰撞 部位 2	超速 1	兩車關係 2	是否過路 口中心 2
屬性值	1,5,7	1,2,7,8	2,5,6	3,4,7	5	1,4	1	2,3,7,8	2
節點	19	20	21	22	23	24	25	26	27
屬性	事故位置 型態 1	事故位置 型態 1	動作 2	幹道或支 道 2	動作 1	事故位置 型態 1	閃光號誌 1	車種 1	車種 1
屬性值	3,5	3,5	3,4,5	2	6,7	1,2	2,3	3	2,4
節點	28	29	30	31	32	33	34	35	36
屬性	是否過路 口中心 1	車種 2	年齡 1	兩車關係 1	車輛碰撞 部位 1	車種 2	超速 1	車輛碰撞 部位 1	超速 1
屬性值	1	2	5	1,8	2,5	4	2	3	2,4
節點	37	38	39	40	41	42	43	44	45
屬性	年齡 2	駕駛人飲 酒 1	兩車關係 1	兩車關係 2	事故位置 型態 1	變換車道 2	車輛碰撞 部位 1	變換車道 1	車輛碰撞 部位 1
屬性值	1	1	2,8	8	1,5	4	4,5	3,4	5
節點	46	47	48	49	50	51	52	53	54
屬性	超速 1	違規行駛 2	車輛碰撞 部位 2	年齡 1	超速 1	車輛碰撞 部位 2	年齡 1	動作 1	車輛碰撞 部位 2
屬性值	1	1,6	2,3	1	4	1,3	4,5	4,7	1,4
節點	55	56	57	備註：節點數字與圖 1 中之代號是相對應的欄位； 屬性欄位中後面的數字 1、2 代表當事者一 或是當事者二；屬性值所對應類別名稱可於 本研究中表 2 中取得。					
屬性	動作 2	車輛碰撞 部位 2	年齡 1						
屬性值	5	2	3,5						

由表 6 可知，最佳模式之末端節點共有 58 個，具有代表性(純度 70%以上)的末端節點有 28 個，而尚具代表性(純度 50%-70%)的末端節點有 15 個，則不具代表性(純度 50%以下)的末端節點有 15 個。

表 6、最佳模式之末端節點屬性與純度之彙整表

節點	58	59	60	61	62	63	64	65	66
標記	無因	無因	完全	無因	完全	完全	主因	主因	主因
純度	66.7%	66.7%	50%	100%	100%	100%	69.2%	90%	100%
節點	67	68	69	70	71	72	73	74	75
標記	完全	次因	次因	主因	主因	次因	主因	同為	主因
純度	42.9%	55.6%	41.7%	78.7%	57.1%	75%	72.7%	75%	50%
節點	76	77	78	79	80	81	82	83	84
標記	次因	主因	無因	次因	無因	主因	無因	完全	主因
純度	80.4%	47.7%	73.7%	43.5%	42.2%	38.5%	66.7%	67.3%	73.9%
節點	85	86	87	88	89	90	91	92	93
標記	完全	主因	完全	主因	次因	次因	次因	完全	無因
純度	51.2%	33.3%	50%	100%	100%	100%	100%	100%	89.7%
節點	94	95	96	97	98	99	100	101	102
標記	無因	主因	完全	無因	主因	完全	同為	完全	完全
純度	73%	33.3%	38.5%	42.9%	75%	93.8%	100%	50%	64.9%
節點	103	104	105	106	107	108	109	110	111
標記	完全	完全	完全	完全	主因	完全	完全	主因	主因
純度	47.1%	45.4%	71.4%	50%	75%	90.6%	97.4%	46.7%	100%
節點	112	113	114	115	備註：標記中「完全」為完全原因；「主因」為主要原因；「同為」為同為原因；「次因」為次要原因；「無因」為無肇事原因。				
標記	完全	無因	完全	次因					
純度	100%	50%	42.9%	41.5%					

陸、結論與建議

6.1 結論

本研究採用決策樹進行肇事事務鑑定系統之預測，所得結論有以下幾點：

- 一、本研究所構建之最佳模式為所有事故類型混合輔以雙方變數構建之模式，配合分支節點限制樣本為 50 時，訓練判中率達 64.77%，與最佳驗證判中率 71.43%，顯示該模式已有較佳的分類與預測功效，故為最佳模式。
- 二、本研究所構建汽車對汽車與非汽車對汽車之事故類別中，分別以 62.55%與 60.28%為該模式中驗證判中率較佳之模式。
- 三、本研究於案件數所構建模式方面，依照三種事故類型構建分類樹皆有將樹之節點越細分，則判中率越精準之趨勢。

6.2 建議

本小節針對本研究可改善部分與後續可進行相關研究之部分進行建議如下：

- 一、在進行相關模式建構上，因肇事資料庫之案件中，屬於同為肇事原因之樣本數不足，導致分類進行完成後，各分支及末端節點較難分類出該類別，故本研究建議需再進行資料庫擴建，讓該類別能充分呈現出屬性，進而減少驗證時之誤判率。
- 二、茲因採用之三種類型樣本數不一，故若把汽車對汽車與非汽車碰撞之類型的分支節點假設與所有事故混合類型相同，導致該樹發展過於精簡與驗證判中

率過低，進而無法充分預測案件之肇責。

- 三、進行分類樹預測乃以精準為優先考量再考慮精簡過程，故進行修剪樹之分支節點時可採更小間距作為修剪準則之方式進行，避免分支節點之樣本一次限制過多，導致分類預測效果不佳；後續研究可以採節點最小限制數為 25，再評估其判中率是否可以提升。

參考文獻

1. 曾憲雄、蔡秀滿、蘇東興、曾秋蓉、王慶堯，資料探勘，旗標出版有限股份公司，94 年 5 月初版。
2. 吳旭智、賴淑貞，資料探勘理論與實務，維科圖書有限公司，初版，民國 90 年。
3. 王秀雯，「應用資料挖掘技術於交通事故傷亡嚴重程度之研究」，國立嘉義大學運輸與物流工程研究所碩士論文，民國 93 年。
4. 陳文杰，「應用資料挖掘技術於高速公路交通肇事次數之研究」，國立嘉義大學運輸與物流工程研究所碩士論文，民國 93 年。
5. 蔡家昌，「應用決策樹歸納法探討台灣行動電話市場區隔」，國立臺北大學統計學研究所碩士論文，民國 90 年。
6. 張希鳳，「資料採掘於精神分裂症之研究」，樹德科技大學資訊管理研究所，民國 90 年。
7. 曾宏平，「運用資料挖掘技術發展建設公司企業知識-以售後服務資料庫為例」，國立高雄第一科技大學營建工程系碩士論文，民國 91 年。
8. 蕭豐達，「應用模糊理論與決策樹法於空載影像建築物分類之探討」，朝陽科技大學環境工程與管理學系碩士論文，民國 93 年。
9. 邱裕鈞、郭曜彰，「汽機車碰撞事故之鑑定決策支援系統」，94 年度車輛行車事故鑑定研究中心成果技術推廣研討會，頁 4、13-4、25，民國 94 年。
10. 邱裕鈞、方守潔，「兩車碰撞事故之肇事鑑定專家系統」，中華民國運輸學會第 19 屆論文研討會論文集，頁 985-998，民國 93 年。
11. 邱裕鈞、王福聖、紀竣傑、郭曜彰、謝侃君，「兩車碰撞事故之鑑定決策支援系統」，93 年道路交通安全與執法國際研討會，頁 341-353，民國 93 年。
12. 葉名山、林世淵，「事故鑑定芻型與理賠機制之研究」，中華民國運輸學會第 19 屆論文研討會，頁 1191-1215，民國 93 年。
13. 楊宗璟、侯至隆、葉筱雯，「直行車與同向右轉車發生碰撞時致使一方無肇事責任之鑑定資料分析」，中華民國運輸學會第 19 屆論文研討會，民國 93 年。
14. 楊宗璟、林芳誼，「路段中變換車道者與同向直行車發生碰撞事故責任與傷亡統計模式之構建」，中華民國運輸學會第 19 屆論文研討會，民國 93 年。
15. 楊宗璟、陳婉郁，「路口橫向碰撞事故責任鑑定之統計分析及模式建構」，中華民國運輸學會第 18 屆論文研討會論文集，民國 92 年。
16. 陳高村、廖信智，「交通事故原因分析鑑定準則之研究」，中華民國運輸學會第 17 屆論文研討會論文集，頁 515-524，民國 91 年。
17. 楊宗璟、艾嘉銘，「交通違規事故責任鑑定影響因素與結果統計分析及模式構建」，中華民國運輸學會第 17 屆論文研討會，民國 91 年。
18. 張漢威，車禍處理與鑑定實務，二版，民國 90 年。
19. 張漢威，車輛肇事鑑定之研究，初版，民國 90 年。

- 20.周文生、陳蔚文，「道路交通事故肇事原因分析程序之研究」，88年道路交通安全與執法研討會，民國88年。
- 21.吳宗修，「非號誌化交岔路口事故責任鑑定原則之探討」，87年道路交通安全與執法研討會，頁279-287，民國87年。
- 22.陳高村，「不同碰撞型態與行為肇事原因分析與責任鑑定」，87年道路交通安全與執法研討會，頁267-278，民國87年。
- 23.陳高村、龍天立，「侵權形態事故肇事責任鑑定方法之建立」，中華民國第二屆運輸安全研討會，頁257-266，民國84年。