

## 以資料探勘技術建構巨觀酒駕肇事預測模型

郭佩棻<sup>1</sup>  
周文生<sup>2</sup>  
徐晨剛<sup>3</sup>  
關迺璇<sup>4</sup>

### 摘 要

酒駕肇事是台灣地區重要的交通肇事主因之一。過去研究多認為酒駕肇事發生之時空特性與飲酒店密度等環境因子高度相關。然而，目前交通肇事預測模型通常使用負二項迴歸式，並以微觀角度(路段或路口為基礎)進行分析;所使用的變項多為交通量或道路與交通特性等;無法將其重要環境影響因子，如:飲酒店密度與土地使用特性納入模型加以預測。另外，國外文獻現有之酒駕肇事預測模式是基於他國之地區交通特性所建議，與台灣土地使用、道路構造、交通及駕駛行為多不相同。

因此，本研究以村鄰里為分析單位，並利用 GIS 技術結合肇事資料、道路交通特性、社會經濟變項與飲酒店資料等多項資料庫，以建立巨觀酒駕肇事預測模型;並使用一般肇事資料作為對照組加以比對其解釋變項之不同。另外，因巨觀酒駕肇事預測模型之模型架構未知且牽涉變數繁多，故本研究建模過程係使用資料探勘採礦技術(分類迴歸樹與隨機森林)，從可能變數中選擇重要之解釋變數而，而非使用傳統的統計建模方式來選擇模型變數。本研究初步結果顯示影響台北市臺北市酒駕肇事之重要預測變數包括:公共運輸服務點數、車流量、犯罪率與 CCTV 個數皆與酒駕肇事數相關。

**關鍵字：**酒駕肇事、資料探勘、飲酒店、肇事預測、隨機森林

### 一、序 言

酒駕肇事是台灣地區重要的交通肇事原因之一。根據警政署統計，近五年酒醉(後)駕車死傷人數(如圖1):在修法之前(2010~2012年)，約有20%是酒駕肇事的死亡車禍，且平均每天會因酒駕造成一件A1類肇事;雖然2013年後酒駕肇事數量漸少，但仍舊佔A1類車禍10%之多。針對前述酒醉駕車不法行為，警政署每月規劃取締酒醉駕車專案勤務，並要求各警察機關針對轄內易肇事或酒醉駕車(如餐飲業、酒店、PUB、KTV等易飲酒場所)之周遭地區、路段與時段，因地制宜規劃勤務部署。根據警政署統計指出:2014年A1類酒醉(後)駕車肇事時間以凌晨時段「0-6時」發生56件(占35.00%)最多，2013年則以晚間時段「18-24時」發生99件(占42.31%)最多，警方認為酒駕與夜間應酬及敬酒文化有關，且其發生地區與

<sup>1</sup> 中央警察大學犯罪防治學系助理教授。

<sup>2</sup> 中央警察大學交通學系暨交通管理研究所教授(聯絡地址:桃園市龜山區大崗里樹人路56號,電話:03-3282321 轉 4513, E-mail:una141@mail.cpu.edu.tw)。

<sup>3</sup> 中央警察大學犯罪防治學系碩士生。

<sup>4</sup> 中央警察大學犯罪防治學系碩士生。

飲酒店空間相關。此主張與其他學者一致，如葉毓蘭等及高鳳仙等人（2015）將酒駕歸於夜店所衍生相關違法案件之一，並主張對大安區、信義區等夜店周邊進行酒駕管理。然而，國內少有相關研究系統性回顧或以實證資料估計此種酒駕專案之績效。

然而，目前交安學術研究常用之肇事迴歸模型，基於美國公路安全手冊（Highway Safety Manual, HSM）建議，一般研究通常是使用負二項迴歸式，並以微觀角度（路段或路口為基礎）進行分析；所使用的變項多半為交通量或道路與交通特性等。若沿用此設定於預估台灣地區的酒駕肇事模型有下列缺點：(1)HSM模式是基於美國地區交通特性，與本國道路構造、交通及駕駛行為不同；(2)單獨使用微觀模型並不適合來預測酒駕事故，因其重要環境影響因子，如：飲酒店密度與土地使用特性皆必須使用巨觀模型加以分析。因此，本研究以村鄰里為分析單位，並利用 GIS 技術匯入肇事資料定義酒駕肇事熱點，並結合道路交通特性、社會經濟變項與飲酒店資料等多項資料庫，建立巨觀酒駕肇事預測模型；另外，使用一般肇事資料作為對照組加以比對其解釋變項之不同。最後，因巨觀酒駕肇事預測模型之架構未知，且牽涉變數繁多，故本研究建模過程係使用資料探勘技術（分類迴歸樹與隨機森林），從可能變數中選擇重要之解釋變數。而非使用傳統的統計建模方式來選擇模型變數。

本文架構為第二節回顧國內外酒駕肇事相關文獻，第三節介紹研究資料與方法，第四節初步研究成果，最後是第五節綜整結論與建議。本文所指酒駕肇事，是指駕駛人在「有飲酒」的前提下，發生「道路交通事故」。換句話說，若「道路交通事故初步分析研判表」當中的「肇事原因研判」註記為「酒（醉）後駕車失控」之事故本研究定為「酒駕肇事」，其他為「一般肇事（非酒駕）」。

## 二、文獻回顧

過去探討資料探勘於交安研究與酒駕肇事因素之相關文獻，主要可概分為五大類：

### 2.1 資料探勘技術應用於交通安全研究

資料探勘技術廣泛應用在交通安全領域；近年來已有許多文獻研究各探勘技巧在定義肇事風險與肇事傷亡程度之重要預測變數、建構非線性肇事預測模型，改善模式預測績效或是利用即時資訊（如車輛感測器或濃霧偵測器）分析肇事風險提出立刻之警告（Chang and Chen, 2005; Chang and Wang, 2006; Montella et al, 2011, 2012; Li et al., 2008; Harb et al., 2009; Pande et al., 2010; Chung, 2013; Wang et al., 2016; Ling, et al., 2015; Wang, et al., 2015; Qi & Abdel-Aty, 2015）。常見之技術包括分類迴歸樹（Chang and Chen, 2005; Pande et al., 2010; Chung, 2013）、隨機森林（Wang et al., 2016）、支援向量機器（Li et al., 2008; Yu & Abdel-Aty, 2013）、類神經網路（Abdel-Aty, 2008）、貝氏網路模型（Hossain & Muromachi, 2012）、關聯法則（Montella et al, 2012）與頻繁模式樹（Lin et al., 2015）等。此類方法雖能即時處理大量且多元之資料，但其複雜之原理與方法也增加交安分析師實際使用的困難度。因此，本研究將於後續方法論章節先介紹最基本但使用頻繁的兩主要資料探勘技術：分類樹與隨機森林。

## 2.2 飲酒與駕駛行為及交通事故的關聯

飲酒會影響人類神經，使駕駛人亢奮、失神睏倦與動作遲鈍，進而影響行車能力造成肇事。早期相關研究如 Borkenstein 等（1964）分析 5,895 個肇事司機和 7,589 個未肇事司機的酒精含量，肇事發生地點與司機之背景資料等，該研究主張車禍風險與駕駛酒精濃度相關（帕翠夏.瓦拉，楊秋和，2000）。陳存雄(2011)彙整飲酒對行車能力之影響。由於酒精會麻痺大腦新皮質反應力、判斷力均降低，意識模糊無法自我控制，危險知覺下降，視覺下降反應遲鈍。導致後果如：方向盤握不穩，行車速度忽快忽慢，蛇行搖擺，或撞上安全島等障礙物，撞衝行人和他車，亂闖紅燈，或是開車途中昏睡等危險駕駛行為。酒精除會使人動作遲鈍，反應減慢而發生車禍，亦會產生「視覺隧道」效應，左右狀況看不清，更難看清前方路況，猶如在隧道中行車，因此對同行車輛及兩側闖入之事物之感應遲鈍，且無法穩定操控車輛，此時容易發生擦撞、衝撞等交通事故。

## 2.3 飲酒店與酒駕肇事之時空分布

國外有許多研究探討環境、飲酒及相關社會(如酒駕肇事)問題。早期研究 Gruenewald 等（1996）指出，販售酒類店面密度高的地方，其鄰近區域的酒駕肇事率也高。該研究將多項變項納入空間分析，如：環境特性（路網密度、車流、人口密度）、社會經濟因素（年齡、性別、種族、婚姻狀況、收入、就業狀況）、居民飲酒狀況（戒酒率、飲酒量及頻率）。結果顯示販售飲酒店面與設施密度（即取得飲酒方便）與自陳的酒駕、醉駕行為無關，但和夜間單一車輛肇事率顯著相關，而鄰近地區的酒駕肇事率也受影響。

Treno 等（2007）之六年縱貫性研究亦顯示：人口與環境特性之變動和車禍有關，飲酒零售處數量與車禍肇事率有關，特別是酒吧。另外 Gruenewald 與 Ponicki(1995)分析美國 38 個州 12 年以上的時間序列橫斷性資料，探討飲酒販售對於夜間單一車輛肇事(single-vehicle nighttime, SVN) 死亡車禍的影響性，結果顯示 SVN 死亡車禍發生率與啤酒的販售有強烈關聯，烈酒和葡萄酒則無顯著關聯。Gruenewald 等人（2002）調查美國加州居民發現，販售酒精店家密度與酒精相關肇事有關，每提升 10%的店家密度會導致 3%的酒駕事件發生。Escobedo 與 Utiz（2002）分析新墨西哥州的資料發現，飲酒店家密度每增加一單位，酒駕肇事加 2.4 單位、酒駕死亡肇事增加 0.22 單位。Stockwell 與 Gruenewald（2001）指出，酒駕與機動車事故都和販售酒精營業場所有關。

亦有相關研究結果呈現相反的結論。例如 Stevenson 等人（1998）以 GIS 探討飲酒相關肇事是否發生在飲酒店附近，其比較 1994-1995 年 17 歲以上「曾發生單一車輛酒駕事故者」和「單一車輛非酒駕事故者」受訪者，發現兩個組別肇事點均約在距離飲酒店外 0.49 英里，認為飲酒店位置與酒駕肇事無顯著關聯。Meliker 等（2004）探討血液酒精含量、交通資料、區域人口普查資料、販酒店家與酒駕肇事關聯，使用 Besag - Newell 空間分析法、傳統邏輯斯。結果顯示都低人口密度區，酒駕肇事會比高於預測，另外透過距離分析、邏輯斯迴歸、卡方檢定呈現販酒店家與酒駕肇事並無顯著關聯。

國內研究方面，吳金白（2011）以次級資料分析法及焦點團體訪談為研究方法，

並透過肇事案件之傳統統計分析，指出酒駕肇事再犯行為人被查獲原因均為肇事後於警察現場時查獲，肇事時間多為夜間至凌晨，飲酒機會與次數愈多，其再犯酒駕肇事之機率就愈高。此外，飲酒處所對酒類無限制地提供，而過量飲酒以致酒後駕車再犯肇事，KTV 小吃部、酒店、PUB、餐廳（飲食店）都是常見的飲酒處所之一。

## 2.4 行車環境與酒駕肇事之時空分布

紐西蘭的 Keall 等人（2005）提到，夜晚比白晝容易發生車禍的可能原因有：晚上飲酒機會高、疲勞駕駛、視線不良等。他們將夜間駕駛的風險區分為二：酒精風險或是夜間固有（inherently）的風險；整體來說，對於週末夜晚駛於低流量路段的 40 歲以下駕駛人，酒後駕車佔一半以上的風險來源，該結果與其他相關研究的結論「酒駕者不走高流量路段」合致。

Ponnada（2012）分析 2007 至 2010 年美國俄亥俄州的交通肇事資料、人口數、登記車輛數量、領有駕照的駕駛人數、日均車輛行駛英里數與各郡的酒駕肇事率，該研究指出 Franklin、Cuyahoga、Hamilton、Summit、Montgomery、Lucas 這些高度都市化、人口稠密、交通流量高的城鎮及其周遭城鎮，都是絕大多數酒駕肇事的聚集地。人口密度、交通量、登記車輛數量、與領有駕照的駕駛人數可幫助定義出酒駕肇事率高的熱區。

Treno 等人（2007）分析六年之間的資料，納入人口資料包括家戶人口數與人口統計資料，再加上零售業營業處所、車禍等歷史資料，進行空間自相關分析，結果顯示人口與環境特性之變動和車禍有關。Gruenewald 等（1996）研究也納入路網密度、車流、人口密度作為環境因素加以分析，此外再加上社會經濟因素（年齡、性別、種族、婚姻狀況、收入、就業狀況）、居民飲酒狀況（戒酒率、飲酒量及頻率）。

Gruenewald 與 Johnson（2010）透過交通流量模式，運用交通流量、店家位址、人口密度、路網密度、夜間單一車輛肇事等資料，輔以電話調查當地居民酒駕人口數，以傳統描述統計、空間自相關、空間落遲迴歸模型、時間序列空間等分析，呈現販酒店家密度、酒駕者密度對於機動車酒駕肇事之影響，發現店家密度與酒駕肇事之關聯受到交通流量的影響，而酒駕人口、店家密度會交互影響而增加 SVN 發生率。另外，該研究亦指出店家與肇事之間的關聯受到交通量干擾的，換句話說是飲店家的影響是透過高交通量的環境，使得酒駕車輛發生酒駕肇事（如圖 2-2-1）。飲酒家對於酒駕肇事的影響可以說與當下背景環境高度依賴的（highly context dependent），聯結了當地的交通量、酒駕駕駛人的源頭、鄰里特性。

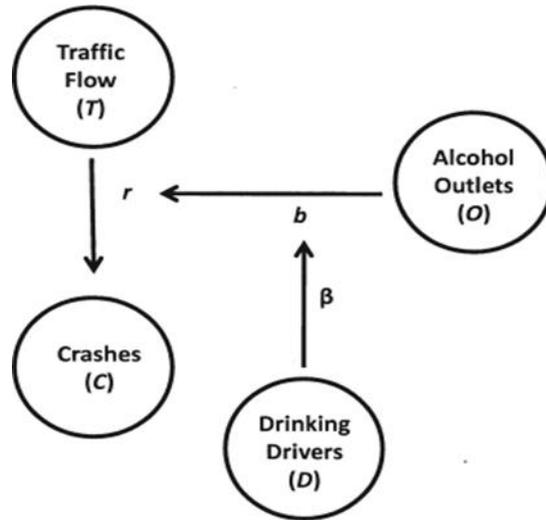


圖 2 飲酒點、車流、酒醉駕駛人與肇事關聯圖 (Gruenewald & Johnson, 2010)

## 2.5 大眾運輸與酒駕肇事之關聯

吳金白 (2011) 指出，酒駕肇事再犯行為人自陳之再犯原因，以僥倖心理最多，其次是風險認知不足，再次則是因為沒有交通工具可搭乘（大眾運輸或他人車輛）、翌日需用車、便利、計程車龐大花費不經濟等考量，加上處於深夜、地點偏僻，仍執意駕車造成不能安全駕駛之風險。王安邦 (2008) 研究指出，酒駕再犯的原因主要是：沒有交通工具可搭乘、很少被警察查獲、雖有喝酒但若自認神智清楚則仍可駕車上路、翌日需用車不得不將車開回家、無適合的大眾交通工具或其他車輛可搭乘、為了省時及方便。

Jackson 和 Owens (2011) 探討華盛頓州交通運輸、飲酒、酒駕間的關聯性，發現如果酒吧周遭有步行可達的大眾運輸系統，則當地酒駕被逮捕的案件會較少，該研究認為：酒吧周圍有公車站 (transit stops) 以及如果深夜還有大眾運輸工具的話，飲酒人回家可使用替代運具，降低自行駕車發生酒駕的風險。

綜上，目前研究對於引起酒駕肇事的相關因子未有定論，且國內相關實證研究較少。因此本研究的主要研究目的是：利用國內資料建構酒駕肇事預測模型，並定義相關環境因子。本研究將飲酒店當作酒駕者主要的飲酒來源，根據 Tobler (1970) 提出地理學第一定律：「每一件事情與其他事情都有關聯性，而距離越近其相關性越高，越遠其相關性則越低。」假設飲酒店周遭有較多的酒後駕車行為，其數目應隨與飲酒店距離增加而遞減；酒客飲酒之後可能會先考量飲酒店周遭的大眾運輸系統是否容易到達、是否營業，以作為其理性決策的參考依據；酒駕車輛行駛在道路上，應會受到當下路段狀況的影響，車流量多、車速快，均可能造成行車的難易度提升。

### 三、研究資料與方法

#### 3.1 研究資料來源

本研究探討酒駕肇事分布與相關環境因素之關聯性，需串聯肇事、道路、交通與社會經濟等多項資料庫。考量各縣市資料完整性與研究成果代表性，本研究選取臺北市為探討對象。台北市自共有十二個行政區，254 里，總面積 271.7997 平方公里，設籍人口 270 萬 3,444 人。本研究資料來源如表一所示：

表 1 本研究使用官方資料之來源

| 名稱           | 資料來源               | 取得資料內容  |
|--------------|--------------------|---|
| 圖層           | 臺北市政府<br>民政局       | 臺北市鄰、里、區界圖（面積、市代碼、區代碼、市名、行政區、座標 TWD97）。           |
| 臺北市<br>街道圖   | 交通部                | 國道、省道（含快速公路）、縣道、鄉道、都市道路（6 米以上）、產業道路及無路名道路等既路。     |
| 飲酒店          | 經濟部商業司             | 商家數、店家地址。   |
| 酒駕肇事         | 刑事局                | 報案地址、時間、內容、車種、肇事者個人基本資料。                          |
| 大眾運輸<br>可近便性 | 臺北市政府<br>捷運工程局     | 臺北都會區大眾捷運系統車站點位圖（車站名稱、車站位置 TWD97）。                |
|              | 臺北大眾捷運<br>股份有限公司   | 臺北市捷運車站出入口（出入口名稱、編號、座標）。                          |
|              | 臺北市<br>公共運輸處       | 臺北市公車站牌位置圖（站牌編號、名稱、座標 TWD67）。                     |
|              | 交通部                | 臺北市高鐵各站位置（名稱、位置）。                                 |
| 行車環境         | 臺北市政府<br>主計處       | 歷年各行政區、里之人口密度。                                    |
|              | 臺北市<br>交通管制工程<br>處 | 臺北市車輛偵測器（VD）資料（各分區內各車流偵測器之平均車速、總和流量、啓始座標點、結束座標點）。 |

另外，本研究所需使用變數其定義如下：(表二為其敘述性統計)

### 3.1.1 肇事資料

臺北市 2012-2014 年約有十萬件交通事故，平均一年有 3 萬多件肇事。表為臺北市政府警察局取締酒駕情形，本研究的酒駕肇事案件數即「移送法辦/肇事」欄位之總和，共有 442 件酒駕肇事，比舉發件數、移送法辦攔檢獲者還要少得多，因此後續分析時考慮將這三年的酒駕肇事位點加總疊合，避免切割成三年分析致使酒駕肇事樣本數過少的困境。2013 年因修法下修酒測值濃度，因民眾心理上開始有所顧慮，導致酒駕行為與酒駕肇事案件的減少，因此數量有下降趨勢。

### 3.1.2 飲酒店

本研究以登記為「飲酒店業」者為研究變項之一，定義採用經濟部的「從事酒精飲料之餐飲服務，但無提供陪酒員之行業。」並以該店面所在地址（經濟部提供）轉換之 TWD97 X、Y 坐標，作為後續分析「飲酒店」定位點資料。根據經濟部工商資料庫之數據，截至 2015 年 8 月，全臺飲酒店業之商家數共有 4,361 家，而單就臺北市則有 1,859 家，約佔全國 42.63%。根據表 3 可以發現，以大安區（491 家）、中山區（412 家）最多，其次則是松山區（214 家）、信義區（215 家），其他行政區則不超過 100 間（表 3）。

### 3.1.3 大眾運輸

泛指所有向大眾開放、並提供運輸服務的交通方式，由通路、交通工具、站點設施等物理要素構成。廣義而言，公共運輸包括民航、鐵路、公路、水運等交通方式；狹義的公共運輸是指城市範圍內定線運營的公共汽車及軌道交通、渡輪、纜車等交通方式。本研究僅納入常見常用的大眾運輸工具：公車、捷運、火車、高鐵，僅納入該站本體所在的位置（經緯度），不加入出入口位點涵蓋臺北市大眾運輸（包含：公車、捷運、火車、高鐵）共 3322 個位點。

### 3.1.4 人口數密度

以內政部統計之 2013 年底臺北市各里設籍之人口數資料，再加上圖層計算出的各里面積，將人口數除以面積得出人口密度，可用以製作各里、各行政區之人口密度面量圖，了解哪些地區人口較稠密或較稀疏。表 3 為 2012-2014 年臺北市各行政區年底之人口密度，大約在每平方公里 9,900 人左右。車流量與車速資料以台北交控中心之數據繪製線圖層。臺北市人口多聚集在市中心地區，但由於是以設籍在當地居民人數加以統計，因此除了周邊偏遠地區或山區人口密度會比較低，商業區、政府辦公地區也顯得人口密度較低。

### 3.1.5 交通量

臺北市之車流量資料依據為臺北市行控中心提供之每五分鐘車流之數量，本研究取得路段（但並非每個路段皆有設置車輛偵測器，僅有主要路段，此為研究限制之一）之起始座標、結束座標，將兩點坐標相連接繪製成線，各里以所直線平均數為平均交通量。

## 3.2 研究工具

因為酒駕肇事巨觀預測模式中，方程式之型態未知且涵蓋變數項目多，本研究將使用分類樹與隨機森林等兩資料探勘技術，分別討論其酒駕肇事與一般肇事之結果與預測模型。下一章節之研究成果，亦會增加使用傳統統計相關分析與回歸式做為檢驗比對參考之用。

### 3.2.1 迴歸分類樹(又稱決策樹)

交通安全研究通常資料庫多元龐大、自變數目眾多，且相互作用與迴歸模式皆未知，符合資料探勘技術適用之需求。首先，決策樹顧名思義是選出重要變數為節點後以樹狀結構，將巨量資料庫切割為數個同質分組。所謂樹狀結構則是以一筆資料從根部的節點（最上方）進入決策樹。在根部，應用一項決策（如年紀是否大於 C1）來決定這筆資料該進入下一層的哪一個子節點(t2 或 t3)。分類過程會一直重複，直到所有資料到達葉部節點（圖形底端，如 t4 與 t5）。此方法多應用於分群，定義影響因子與建立預測模型。以下概述其原理與應用 (Breiman et al., 1984)。

以圖一為例，若年紀與教育程度會影響駕駛人發生車禍風險高低。則利用年紀和教育程度將母體切割成 t2、t4 與 t5，各子集合內肇事風險應較接近。即肇事者與肇事者（方形）接近，非肇事者與非肇事者（星型）接近。通常使用 Gini 值來判斷子集合中的同質性 (Han et al. 2006)。亦可將 Gini 值視為一純度指標，Gini 值越低則代表純度越高，其計算公式如(1)。今若投擲一正常與”作弊”的硬幣，如投擲”作弊”的硬幣永遠都產生人頭，不會出現數字;而投擲”正常”的硬幣一半機率產生人頭，一半機率出現數字。”作弊”硬幣的 Gini 指數為零，遠低於正常硬幣的 Gini 值 (=詳下方算式)。換句話說，作弊硬幣的投擲結果恆為人頭，純度較正常硬幣一半為人頭一半為字高。

$$i(p) = \sum_{i \neq j} p_i p_j = 1 - \sum_k p_k^2 \quad (1)$$

$$i(p_{unfair}) = 1 - 1^2 - 0^2 = 0 < i(p_{fair}) = 1 - 0.5^2 - 0.5^2 = 0.5$$

$$\text{當 } p_{unfair,head} = 1, p_{unfair,tail} = 0, p_{fair,head} = p_{fair,tail} = 0.5$$

因為決策樹的目的是求同一子集合的資料最相似，因此決策樹的節點會選擇能將變異數降低最多(提高純度)的變數 a。IG 為 information gain 之縮寫，i 為資料庫 T 數值為 x 時的資料純度。關於詳細迴歸分類樹的細節說明，可詳 Chang and Wang (2006) and Harb et al. (2009).

$$IG(T, a) = i(x) - i(x|a) \quad (2)$$

此方法除建立迴歸模型外，亦可依重要性選擇預測變數。因為分類樹可以提高運算效率與正確性，並具備操作簡易、容易理解等特性，在許多領域皆有應用研究，如應用分類樹於預測交通旅次需求與建立交通肇事模型等研究 (Kuo and Shen, 2015; Abdel-Aty et al., 2014)。

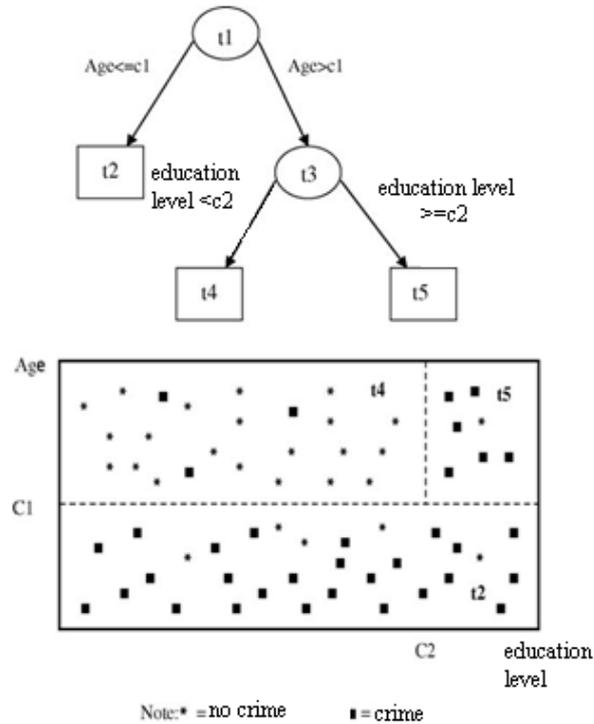


圖 3 分類樹原理示意圖 (Chang & Chen, 2006)

### 3.2.2 隨機森林

隨機森林以分類樹為基礎但較進階的方法。森林一詞指此方法可利用抽樣的概念隨機抽取樣本與變數同時建立大量分類樹。相對於 CART 分類樹而言，隨機森林因為可將變數降維，其處理缺失值、異常值以及不平衡數據時也更有效率及準確。舉例來說，如果 CART 分類樹是我們向單一學者詳細詢問肇事原因所有相關問題，隨機森林就像是同時向多位交安專家抽樣詢問部分問題，但最後綜整其看法，所以其結論較為客觀。另外，因為隨機森林的最後結果需要包括所有分類樹結果，所以節點變項若是連續數值：如年紀，則會以平均值為結果。但若變數是類別變數，則會以多數決定。此處變數與資料抽樣的方式為統計中的拔靴法“bootstrap aggregation”。因其解釋原理可詳 Han et al.(2006)一書。

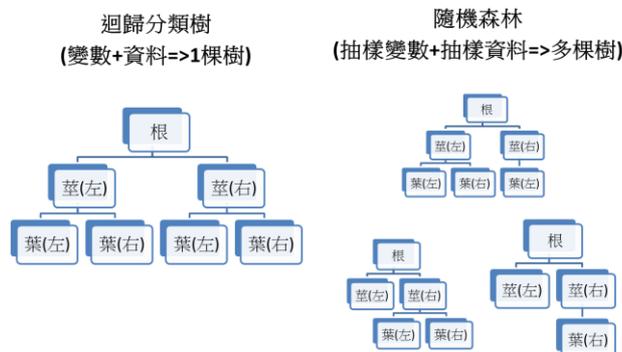


圖 4 隨機森林與分類樹比較示意圖

## 四、結果分析

本章節就前述分類樹與隨機森林等兩資料探勘技術，分別討論其酒駕肇事與一般肇事之結果與預測模型。但亦會使用傳統統計相關分析與回歸式做為檢驗比對參考之用。

### 4.1 相關分析

為瞭解交通肇事發生件數與相關變項（均為連續變項）之間相關程度的高低，本研究以相關分析計算 Pearson 相關係數之方式，呈現 2012-2014 年臺北市 456 個里肇事發生件數與有關變項之相關強度與其顯著性。

表 4 交通肇事數與相關變項之相關分析

|        | 飲酒店數    | 大眾運輸數    | 車流量    | 人口數   | 酒駕肇事數   | 非酒駕肇事數  | 肇事總數 |
|--------|---------|----------|--------|-------|---------|---------|------|
| 飲酒店數   | 1       |          |        |       |         |         |      |
| 大眾運輸數  | 0.014   | 1        |        |       |         |         |      |
| 車流量    | 0.138** | -0.176** | 1      |       |         |         |      |
| 人口數    | -0.003  | -0.122** | -0.069 | 1     |         |         |      |
| 酒駕肇事數  | 0.103*  | 0.217**  | -0.022 | 0.057 | 1       |         |      |
| 非酒駕肇事數 | 0.499** | 0.317**  | 0.064  | 0.064 | 0.435** | 1       |      |
| 肇事總數   | 0.498** | 0.317**  | 0.064  | 0.064 | 0.439** | 1.000** | 1    |

說明：1. \*\*為.001 顯著水準。2. \* .05 顯著水準。3. N=456 (里) 4. 扣除無車流量資料的 99 個里，車流量與飲酒店、大眾運輸、人口、酒駕數、非酒駕數、肇事總數之相關係數及顯著性是以 N=357 個里的資料計算出來。

由表 4 之相關矩陣可以發現：「飲酒店」與車流量、酒駕肇事、非酒駕肇事、肇事總數呈顯著的正相關。結果顯示各里之飲酒店愈多，車流量、酒駕肇事、非酒駕肇事、肇事總數也愈多。與多數研究及預期的相符 (Kumar & Waylor, 2003; Weitzman et al., 2003; Gruenewald et al., 1996; Treno 2007; Gruenewald & Ponicki, 1995; Gruenewald et al., 2002; Escobedo & Utiz, 2002)，與少部分研究不一致，(Stevenson et al., 1998; Meliker et al., 2004)。

「大眾運輸」與酒駕肇事( $r=0.217$ ,  $p<.001$ )、非酒駕肇事( $r=0.317$ ,  $p<.001$ )、肇事總數( $r=0.317$ ,  $p<.001$ )呈顯著的正相關，結果顯示各里之大眾運輸愈多，酒駕肇事、非酒駕肇事、肇事總數也愈多，此與相關研究不一致 (Jackson & Owens, 2011)；此外「大眾運輸」也與車流量( $r=-0.176$ ,  $p<.001$ )、人口數( $r=-0.122$ ,  $p<.001$ )呈顯著的負相關，結果顯示各里之大眾運輸愈多，車流量、人口數愈少。

「酒駕肇事」與非酒駕肇事( $r=0.435$ ,  $p<.001$ )、肇事總數( $r=0.439$ ,  $p<.001$ )呈顯著的正相關。結果顯示各里之酒駕肇事愈多，非酒駕肇事、肇事總數也愈多。「非酒駕肇事」與肇事總數( $r=1.000$ ,  $p<.001$ )呈顯著的正相關。結果顯

示各里之非酒駕肇事愈多，肇事總數也愈多。以上均符合一般的預期。

## 4.2 迴歸分析

表 5 將酒駕與非酒駕肇事分別進行傳統迴歸分析，從自變項與依變項之間之關係中發現，飲酒店變項之顯著情形不一致，車流量、人口數均未達顯著，大眾運輸均達顯著水準，機率值皆小於.001，與依變數皆為正向關係，顯示大眾運輸數量愈多，酒駕或非酒駕肇事數量愈多。本研究酒駕與非酒駕肇事之解釋變異量  $R^2$  值分別為.082 及.398；至於相關研究，Gruenewald 與 Johnson(2010)之研究為.83，Jackson 與 Owens(2011)之研究酒駕肇事在.255~.356，非酒駕肇事在.166~.287，本研究所得出之數值相對較低。

誤差常態檢定部分採用 Jarque-Bera 指標，結果顯示酒駕肇事模型機率值小於.001，誤差非常態分布。

由此可知。酒駕肇事資料之誤差顯然不是隨機獨立，彼此不獨立、變異數不等值，且非常態分布，顯示 OLS 迴歸解釋酒駕肇事是有瑕疵的。事實上，進行以空間作為分析單位的迴歸，通常會產生誤差，主要原因是觀察值並非抽樣而來，其本身就是母體，觀察值與觀察值之間很難完全獨立，例如交通肇事本身可能會因鄰近場遇而有影響，相鄰地區的交通肇事發生都可能相關。

表 5 酒駕與非酒駕肇事數量與相關變項之 OLS 迴歸分析

| 自變項                    | 酒駕肇事                      | 非酒駕肇事                       |
|------------------------|---------------------------|-----------------------------|
| Intercept              | 0.026961<br>(0.130110)    | -9.085768<br>(-0.226725)    |
| 大眾運輸                   | 0.042364***<br>(5.463513) | 15.470216***<br>(10.316500) |
| 車流量                    | 0.000482<br>(0.458486)    | 0.399252<br>(1.965631)      |
| 飲酒店                    | 0.009452<br>(1.494578)    | 13.181740***<br>(10.777767) |
| 人口數                    | 0.000049<br>(1.852494)    | 0.008967<br>(1.752541)      |
| AdjR <sup>2</sup>      | 0.082132                  | 0.397718                    |
| F-statistic            | 8.963834***               | 59.771328                   |
| N 樣本數                  | 357                       | 357                         |
| AIC                    | 1000.870271               | 4759.879702                 |
| 誤差常態檢定<br>Jarque-Bera  | 215.179769***             | 7912.297354                 |
| 誤差異質檢定<br>Koenker (BP) | 20.923412***              | 85.178748                   |

說明：表格內數字為係數值 Coefficient，括號內為 t 值(t-Statistic)；  
p<.001\*\*\* p<.05\*\* p<.01\*

### 4.3 分類樹結果

在含酒駕肇事與必要預測參數的 456 個村里中，本研究利用分類樹模型計算的結果如圖 5 所示。最上方的第一節點為大眾運輸的數量，指該村鄰里中公車、捷運、火車、高鐵等場站或站牌的數量。由分析可知，大眾運輸的場站與站牌數量，若里內大眾運輸服務點大於 10 個，則酒駕肇事較高，推論因為在大眾運輸服務較發達處，應為市中心商業區且居民旅運需求高，因此車禍風險亦增大。其次，第五節點則是里內死亡人數，里內大於 54 人死亡可作為指標；如前面節點的分析所知，在交通運輸量大的里內，酒駕數量多，也能直觀反映在死亡人數上。監視器也是影響酒駕肇事因素之一，第三節點中監視器高於 66 個之里有較高肇事，因為駕駛者倘若不遵守交通規則有超速等危險駕駛行為增加肇事風險，容易被監視器記錄。而在第 11 與 20 節點中，當道路密度小於 9.3 與 9 單位時，較容易發生肇事次數，顯示在相同總面積相同的比較上，里內道路長度會影響肇事，可能原因為因為道路密度為道路總長度與總面積的比值，越小的密度表示里內巷弄、一般道路較少，且有較多能使駕駛者高速馳騁的快速道路或高速公路，致使肇事率高。在第 21、42 中，低教育人口數與結婚數會影響肇事，推估可能原因為在低教育人口數低與結婚數高的里內，因為家庭結構的需求，婚後小家庭持有汽車的比率較高，而在教育程度影響社經地位下，較高教育程度持有汽車率高，因而在家庭、工作需求下時常駕駛，因此肇事率亦高。

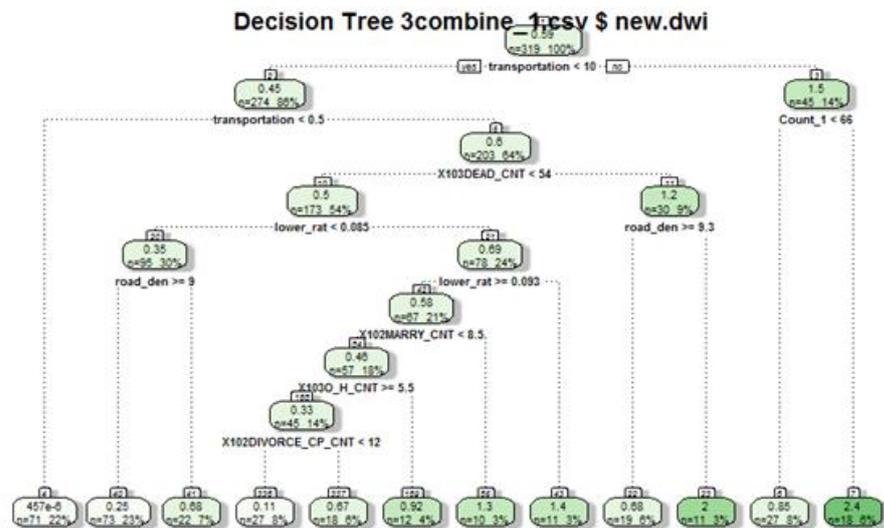


圖 5 分類樹分析結果(酒駕肇事)

在一般肇事（非酒駕）部分，分布於 456 里共計 92002 筆肇事資料。分類樹模型結果如圖 6。第一節點為大眾運輸的場站與站牌數量，若里內大眾運輸服務點大於 7.5 個；則一般肇事較高，推論因為在大眾運輸服務較發達處，應為市中心商業區且居民旅運需求高，因此車禍風險亦增大。第二個節點亦為大眾運輸密度，當大眾運輸服務點數減至 0.5（即里內幾乎無大眾運輸系統），該處應為偏僻山區（如北投士林等地）當地居民旅運需求較低故車禍風險較低。第三、四與五節點為犯罪率高於 0.02 處，酒吧密度高於 5.5 處，

與監視器個數多於 44 個處，其一般肇事次數亦較高，此三節點可利用環境犯罪學日常活動理論解釋：監視器一般多裝在犯罪率較高地區。該地區若犯罪率高且有犯罪吸引處（酒吧），駕駛者容易不遵守交通規則常有超速等危險駕駛行為增加肇事風險。此外，在第十節點中，交流道的密度與道路密度也與一般肇事有關，在交流道高的里內發生肇事的數量較多，交流道多表示高速公路多且車流匯合處多，衝突點多發生風險機率也較高。最後，在第十三節點中，道路密度低一般肇事高，在道路密度小於 7.1 公里時，發生一般肇事的數量較多，可能是因為道路密度為道路總長度與總面積的比值，道路密度越大，表示小巷弄與一般道路多，高速公路較少，在考量駕駛速度下，一般道路上發生車禍的機會較低。

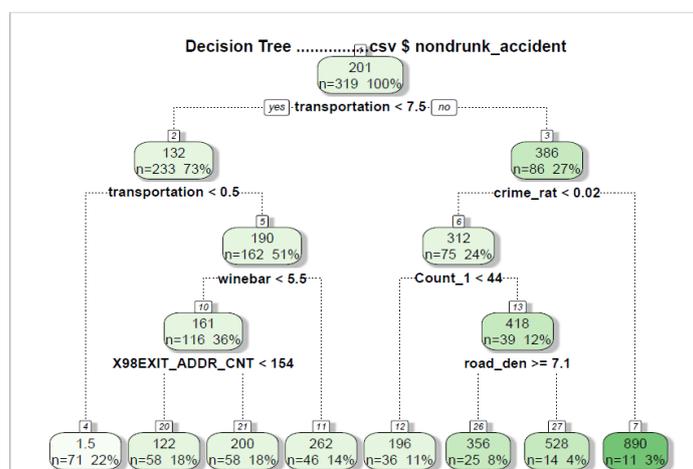


圖 6 分類樹分析結果（未酒駕肇事）

#### 4.4 隨機森林結果

首先，在酒駕肇事部分，以隨機森林決定重要預測變數為：公共運輸數量與車流量，犯罪率與 CCTV 個數。圖 7 是以預測正確率定義之結果圖右是以最大均質來定義。該結果與前述分類樹結果節點定義之重要參數一致，顯示隨機森林根據分類純度之結果，大眾運輸站牌數量、平均車流量、犯罪率與監視器數量皆影響酒駕肇事。

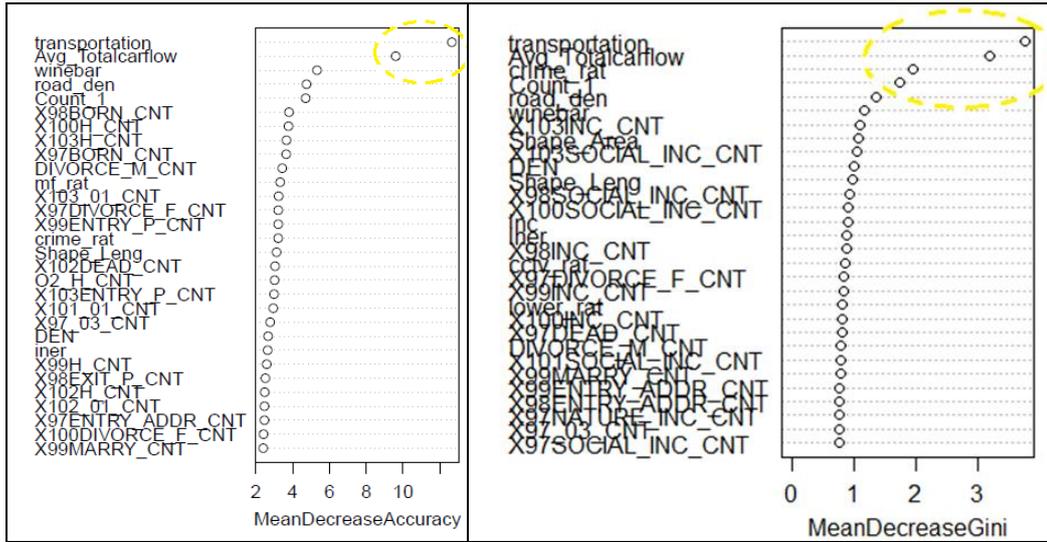


圖 7 隨機森林：酒駕肇事之重要預測變數（依最小誤差原則）

另外在未酒駕而肇事中，下圖 8 結果顯示根據預測準確性，平均車流量、大眾運輸站牌數量、監視器數量、飲酒店數量、犯罪率與未酒駕而肇事有影響。根據下圖 8 隨機森林分類純度之結果，監視器數量、大眾運輸數量、飲酒店數量、犯罪率、平均車流量與未酒駕而肇事有影響。

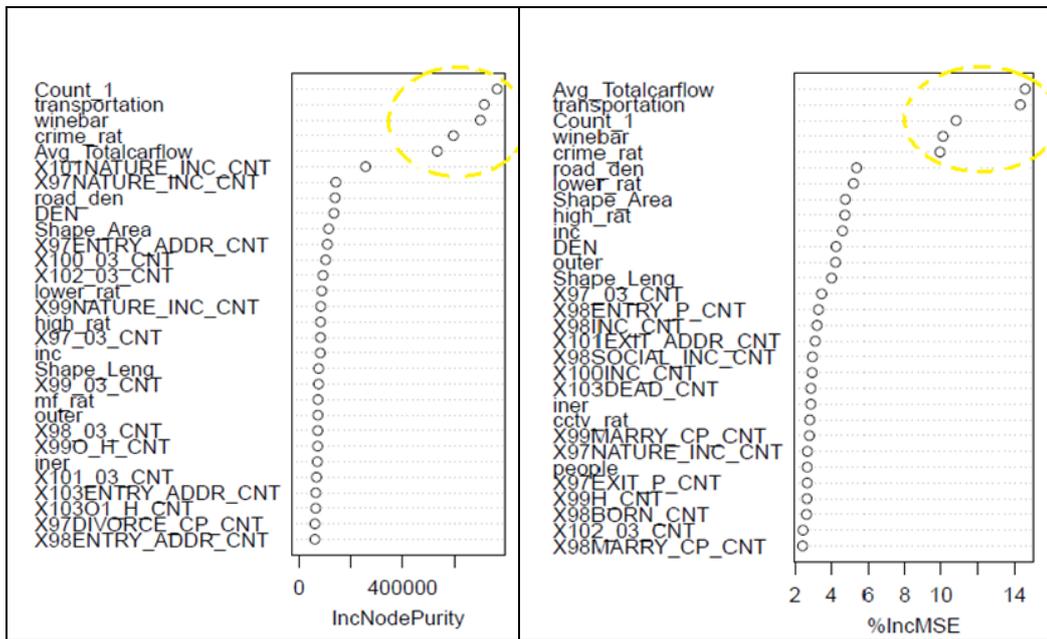


圖 8 隨機森林：未酒駕肇事之重要預測變數（依最小誤差原則）

## 五、結論與建議

本研究透過資料探勘方法，針對台北市酒駕肇事資料，定義其重要影響變數並建立巨觀預測模型。結果顯示：公共運輸服務點數、車流量、犯罪率與 CCTV 個數皆與酒駕肇事數相關。若比較一般肇事資料，結果顯示預測變數類似。值得注意的是里內飲酒店數量與酒駕肇事無顯著相關，但與一般肇事相關。

本研究將初步成果詢問執行酒駕警方建議推論，其可能原因如下：1.酒駕駕駛可由其他管道取得酒精，如小吃店、便利商店或餐廳（非經濟部註冊之飲酒店業）。國內對銷售酒精場所之管制不比國外嚴格，因此研究結果並未如國外文獻飲酒店密度與酒駕有明顯之正相關；2.本研究空間分析單位為里，酒駕肇事駕駛人可能由 A 里飲酒後，離開駕駛一段時間後至 B 里並發生車禍。因此飲酒店反而與附近鄰里之酒駕數相關。3.多數酒駕肇事可能為輕微財損為肇事黑數，並未涵蓋於本研究資料內，建議未來可於量化基礎上融入質性訪談，探究酒駕肇事成因；另有鑒於酒駕肇事較少，後續可擴大研究資料與方位以作出更完整之結論建議。

最後，資料探勘雖可有效分析巨量資料，並提供準確預測但其解釋度往往不盡理想(Tollenaar1 &Heijden, 2012)。有別於商業應用主要訴求為及時預測提高銷售，對於警政單位而言，該如何解釋其變項並擬定有效可行之對策方是首要考量。建議後續研究可針對定義之顯著變數與其門檻值進行更進階之研究，分析結果也需以交通安全學理分析判斷後，再提出經學理驗證之可信策略，精進執法效率，有效協助各項警政方案推動。

## 參考文獻

- 王安邦 (2008)。酒醉駕車決意歷程與預防對策之研究-以高雄地區為例，國立中正大學犯罪防治研究所碩士論文。
- 吳金白 (2011)。酒後駕車肇事再犯危險因子之研究，國立中正大學犯罪防治研究所碩士論文。
- 帕翠夏. 瓦拉、楊秋和 (2000)。與酒精有關的車禍傳染病學與大洪流域研究，警學叢刊，30:4，頁133-152。
- 高鳳仙、孫大川、江綺雯、林雅鋒、陳慶財 (2015)。夜店（飲酒店）違法及犯罪防治專案調查研究，監察院。
- 陳存雄 (2011)。酒後駕車駕駛行為的反應與防制酒後駕車策略之探討，100年道路交通安全與執法研討會。
- Borkenstein, Crowther, Shumate, Ziel & Zylman (1964). The role of the drinking driver in traffic crashes. Bloomington, Indiana, USA: Department of Police

Administration, Indiana University.

- Chang, L.Y., Chen, W.C., 2005. Data mining of tree-based models to analyze freeway accident frequency. *J. Saf. Res.* 36 (4), 365–375.
- Chang, L.Y., Wang, H.W., 2006. Analysis of traffic injury severity: an application of non-parametric classification tree techniques. *Accid.Anal. Prev.* 38 (5), 1019–1027.
- Escobedo & Ortiz ( 2002 ) . The relationship between liquor outlet density and injury and violence in New Mexico. *Accident analysis & prevention*, 34 ( 5 ), 689-694.
- Gruenewald & Ponicki ( 1995 ) . The relationship of the retail availability of alcohol and alcohol sales to alcohol-related traffic crashes. *Accident Analysis & Prevention*, 27 ( 2 ) , 249-259.
- Gruenewald, Remer & Lipton ( 2002 ) . Evaluating the alcohol environment: community geography and alcohol problems. *Alcohol Research and Health*, 26 ( 1 ) , 42-48.
- Gruenewald, Johnson & Treno( 2002 ) . Outlets, drinking and driving: a multilevel analysis of availability. *Journal of Studies of Alcohol*, 63 ( 4 ) : 460–8.
- Gruenewald & Johnson ( 2010 ) . Drinking, driving, and crashing: a traffic-flow model of alcohol-related motor vehicle accidents. *Journal of studies on alcohol and drugs*, 71 ( 2 ) , 237.
- Harb, R., Yan, X., Radwan, E., Su, X., 2009. Exploring precrash maneuvers using classification trees and random forests. *Accid.Anal. Prev.* 41 (1), 98–107.
- Jackson, & Owens ( 2011 ) . One for the road: public transportation, alcohol consumption, and intoxicated driving. *Journal of Public Economics*, 95 ( 1 ) , 106-121.
- Keall, Frith & Patterson ( 2005 ) . The contribution of alcohol to night time crash risk and other risks of night driving. *Accident Analysis & Prevention*, 37( 5 ) , 816-824.