

逢甲大學  
交通工程與管理學系碩士班  
碩士論文

路口交通事故成因分析方法之比較研究

Applicability of Analytical Methods on  
Analysis of Intersectional Traffic Accident

指導教授：劉 霈

研 究 生：楊仁維

中 華 民 國 九 十 五 年 七 月

## 誌謝

校園的風 依然吹著 操場的樹 仍然擺著  
教室的燈 不滅的亮著 黑板上的粉筆灰 還存在著  
老師的笑容 在我心中保留著  
捨不得離開的家 我 必須走了  
再見了  
我的眼淚 正在悄悄的流著

研究所二年的時光來去匆匆，學習過程饒富趣味並且充滿挑戰，回想六年前剛踏入逢甲，彷彿就像昨日般清晰可見。便當街、丘逢甲紀念館與資電館，陪我一起度過，班上同學陪我一起成長。

兩年學業可以如期完成，首先要感謝指導教授劉霈老師在研究所二年的循循善誘，在我思考困頓的時候有如一盞明燈，不斷地指引我抽絲剝繭終至豁然開朗，謹此衷心誌謝。

論文口試期間承蒙楊明德教授以及陳世晃教授細心審閱，不吝指正，並提供寶貴意見，使本論文更臻完善，在此特表謝意。在受業期間也承蒙邱裕鈞教授於研究方法知識上之啟發，使學生受益匪淺，同時也感謝資料探勘的啟蒙老師 林明言教授明確教導軟體的操作，不斷與老師再討論、再思考，在潛移默化中也增進了自己的能力與自信。特此表示最深感謝。

另外感謝研究所同學，老賴、鵬升、勳哥、怡霖、欣翰、阿展、阿舜、米奇、阿彰、耿禎、公狗、姿雁、阿裘、芋頤、妍菁、A超、憲哥、媽祖、庭銀、小碧、智淮及小雅等，以及資管中心的督導們與全體工讀生夥伴，還有文彥、豪君、淑敏姐及小映學妹，甚至其他好友們，感謝這幾年來你們的支持與關懷，都將使我畢生難忘。

驀然回首，感謝我所有最親愛的家人，一直在背後默默地支持與鼓勵，使我於求學過程能無後顧之憂，謹以此論文表達我最真摯的感謝。

楊仁維 謹誌於

逢甲大學交通工程與管理學系碩士班

## 摘要

在交通事故分析或鑑定結果中，人為因素多被視為交通事故發生的主要原因，且通常為唯一被考量的因素；事實上，交通事故與當事人駛近事故地點時所面臨的道路環境亦有一定程度之相關性，如若路口幾何設計或規劃不當而形成潛在危險因子，即容易導致交通事故發生；而相較於其他交通事故點，交叉路口為最常發生事故的位置。因此，一味地將事故責任推至用路人而不思改善道路環境，不僅不公平，而且無法避面類似事故的重演。若能瞭解路口環境因子對肇事的影響程度並加以改善，或許能將肇事發生的次數與嚴重程度有效的降低。其次，此等資訊亦能提供交通及警察相關單位作為改善易肇事地點安全或評估交通管理措施可行性之參考。

交通事故與駕駛行為、幾何、號誌及交通現況等因素交互影響且錯綜複雜，而各因素對事故之影響程度亦不易界定，因此本研究以統計相關性分析先行篩減模組變數，增加變數代表性及模式可用性。蒐集台中市於民國 88-89 及 92-93 年之路口事故資料，將其分解為 1,039 組行向組合，以類神經網路與負二項迴歸方法進行路口交通事故數量預測模式之構建。結果以類神經網路之預測能力較佳。並再利用敏感度分析及彈性分析篩選出顯著變數。

最後本研究觀察現行作業，對於肇事路口主要肇因之判定，並無一套完整之評估規則，故利用分類與迴歸樹產生較佳之規則判定。結果顯示以 Gini 決策樹之 47 條規則判斷較佳，然各規則於相異之行向組合中採用之準則皆不同，而整體準則之順序為道路寬度、行向流量、機車比、左/右轉車輛比、路口總交通量、快慢車分隔形式、速限或警告標誌與時制週期等。以道路寬度為例，其門檻值為 14.25 米，對應道路設計型態約為單向四車道，並經由實例證明，五權西路寬度為 13.3 米與 12.4 米，為單向三車道(兩快車道一慢車道)，環中路寬度為 26.5 米與 26.2 米，為單向四車道(三快車道一側車道)，並以決策樹規則去判斷，結果顯示危險程度以環中路較為危險，顯見應用 Gini 決策樹於交通領域上有其相當合理性，故應在道路設計考量及改善路口評鑑時具應用之可行性。

關鍵字：路口交通事故、類神經網路、負二項迴歸、分類與迴歸樹

## ABSTRACT

*Most traffic accident analyses and authentication results consider human errors as the main factor, or even the only factor, which induced the occurrence of traffic accidents. However, in fact, the approaching roadside environment that drivers confront right before crash can sometimes play more important role than human error on the occurrence of the crash. Inappropriate geometry design, poor pavement skid resistance, and inadequate sign and signal arrangement can all create a dangerous driving environment. Consequently, putting all the blames to drivers is unfair and similar accidents may occur one after one. Apparently, there is a need to study roadside environment. On the other hand, due to the fact that great amount of traffic accidents occurred at intersections, it is thus worthwhile to study the correlation between nonhuman factors around intersections and traffic accident occurrence. Such information may help agencies to figure out what needs to be done and what should be done first. It may also serve as a reference for future intersection design and management.*

*Factors related to traffic accidents are generally inter-correlated, multivariate regression was thus adopted to analyze traffic accident problems. Yang used too many factors in that model, which made that model difficult to use. In this study, 18 out of the 53 factors used in Yang's model were first sieved out by using statistical dependence test, chose one hundred and two intersections in Taichung City, divided them into one thousand and thirty-nine moving direction combinations, and constructed the forecast model of the number of the intersection traffic accidents with the data of 3,441 traffic casualties happening within four years.*

*Therefore, the study adopted the methods of the Neural Networks and the negative binomial regression. The outcome was that the Neural Networks had the better forecast ability. The correlation coefficient was 0.994; the misjudging rate was 13.53%; the maximum negligence was 1.37; the networks'*

*MSE value was  $2.98 \times 10^{-6}$ . Prediction results of these models were not all very accurate, yet can be considered satisfactory.*

*Finally, 10 out of the 18 factors were sieved out using sensitivity analysis and elasticity analysis. The study used the method of the classification and regression trees to determine intersections dangerous judgement. The outcome was that the Gini tree better than Twoing tree. Being found by the importance variable, the geometry was more important than traffic signs and traffic movement. The order were total road width, moving flow, motorcycle rate, turn left/right vehicle rate, total intersection volume of traffic, lane partition, speed limit or warning sign and cycle.*

Keywords : intersection traffic accident, artificial Neural Networks, negative binomial regression, classification and regression trees



## 目錄

誌謝 .....	I
摘要 .....	II
ABSTRACT .....	III
目錄 .....	V
圖目錄 .....	VIII
表目錄 .....	IX
第一章 緒論 .....	1
1.1 研究背景與動機 .....	1
1.2 研究目的 .....	3
1.3 研究範圍 .....	4
1.4 研究內容與流程 .....	5
第二章 文獻回顧 .....	9
2.1 傳統模式分析交通事故之嚴重程度 .....	9
2.2 類神經網路 .....	11
2.2.1 交通運輸課題之應用 .....	12
2.2.2 其他 .....	14
2.3 資料探勘部分 .....	15
2.3.1 交通運輸課題之應用 .....	16
2.3.2 其他 .....	17
2.4 文獻回顧小結 .....	20
第三章 研究方法 .....	27
3.1 統計方法 .....	27
3.1.1 卜瓦松迴歸模式 .....	27
3.1.2 負二項迴歸模式 .....	29
3.1.3 變數之彈性 .....	32
3.2 類神經網路 .....	32
3.2.1 類神經網路基本概念 .....	32
3.2.2 倒傳遞網路 .....	34

3.2.3 效能評估 .....	37
3.2.4 敏感度分析 .....	38
3.3 資料探勘 .....	38
3.3.1 資料探勘概念 .....	39
3.3.2 分類與迴歸樹 .....	44
3.3.3 Gini分類法 .....	46
3.3.4 Twoing分類法 .....	47
3.3.5 重要性變數 .....	47
3.4 各模式之優缺點比較 .....	48
3.5 小結 .....	51
第四章 變數選取與資料蒐集分析 .....	53
4.1 行向組合 .....	53
4.2 共線性與變數選取 .....	55
4.2.1 共線性之概念 .....	55
4.2.2 變數選取 .....	56
4.3 相關變數之說明 .....	58
4.3.1 工程因素變數 .....	58
4.3.2 交通動態變數 .....	61
4.4 資料之蒐集及初步整理 .....	62
4.4.1 路口及肇事相關資料之蒐集 .....	62
4.4.2 肇事資料之蒐集 .....	64
4.4.3 交通動態資料之蒐集 .....	65
4.4.4 道路環境資料之蒐集 .....	65
4.5 路口交通事故資料初步分析 .....	65
4.5.1 路口事故之次數 .....	66
4.5.2 相異時間之路口事故次數 .....	68
4.5.3 行向組合交通事故次數 .....	73
4.5.4 路口肇事時段 .....	75
第五章 模式建構與分析 .....	77
5.1 敘述性統計分析 .....	77
5.2 統計模式與分析 .....	78

5.2.1 負二項模式建立 .....	78
5.2.2 彈性分析 .....	83
5.3 類神經網路模式與分析 .....	84
5.3.1 類神經網路模式建立 .....	84
5.3.2 敏感度分析 .....	90
5.4 負二項模式與類神經網路模式比較 .....	93
5.4.1 預測正確率比較 .....	93
5.4.2 敏感度分析與彈性之比較 .....	94
5.5 小結 .....	95
第六章 道路危險性指標與決策樹分析 .....	96
6.1 道路潛在危險性評估指標 .....	96
6.2 模式建構 .....	99
6.3 分類與迴歸樹建立與分析 .....	100
6.3.1 Gini結果 .....	101
6.3.2 Twoing結果 .....	103
6.3.3 預測正確率比較 .....	104
6.3.4 變數重要度 .....	106
6.4 Gini決策樹實證分析 .....	107
6.4.1 Gini決策樹節點分析 .....	107
6.4.2 實例測試 .....	115
6.5 小結 .....	121
第七章 結論與建議 .....	122
7.1 結論 .....	122
7.2 建議 .....	123
7.3 研究貢獻 .....	124
參考文獻 .....	125
中文部份 .....	125
英文部分 .....	126
附錄 .....	
附錄A Gini決策樹 .....	附錄A-1
附錄B Twoing決策樹 .....	附錄B-1



## 圖目錄

圖 1.1 人車路因素交集圖 .....	3
圖 1.2 路口範圍平面圖 .....	5
圖 3.1 資料探勘程序的分類 .....	43
圖 3.2 資料探勘技術的分類 .....	43
圖 4.1 四岔路口之行向組合之構成方式 .....	54
圖 4.2 資料整理流程圖 .....	65
圖 4.3 88、89 年路口之交通事故次數分佈圖 .....	67
圖 4.4 92、93 年路口之交通事故次數分佈圖 .....	68
圖 4.5 台中港與文心路交叉路口示意圖 .....	69
圖 4.6 忠明南路與南屯路交叉路口示意圖 .....	71
圖 4.7 行向組合肇事次數分佈圖 .....	74
圖 4.8 選取路口肇事時段分佈圖 .....	75
圖 5.1 負二項迴歸模式預測值與實際值之差值圖 .....	82
圖 5.2 負二項迴歸模式預測圖 .....	83
圖 5.3 類神網路模式模擬訓練資料預測值與實際值之差值圖 ..	88
圖 5.4 類神經網路模式模擬訓練資料預測圖 .....	89
圖 5.5 類神網路模式模擬驗試資料預測值與實際值之差值圖 ..	89
圖 5.6 類神經網路測試資料預測圖 .....	90
圖 5.7 左轉專用道之敏感度分佈圖 .....	92
圖 6.1 肇事危險程度等級 .....	100
圖 6.2 Gini決策樹 .....	102
圖 6.3 Twoing決策樹.....	103
圖 6.3 Twoing決策樹.....	103
圖 6.4 五權西路與環中路交叉路口示意圖 .....	115

## 表目錄

表 1.1 台灣地區機動車輛與道路里程統計 .....	1
表 1.2 台灣地區歷年道路交通事故統計 .....	1
表 2.2 文獻彙整表(一)交通事故分析類 .....	20
表 2.2 文獻彙整表(一)交通事故分析類(續) .....	21
表 2.3 文獻彙整表(二)類神經網路應用類 .....	22
表 2.3 文獻彙整表(二)類神經網路應用類(續) .....	23
表 2.3 文獻彙整表(二)類神經網路應用類(續) .....	24
表 2.4 文獻彙整表(三)資料探勘應用類 .....	24
表 2.4 文獻彙整表(三)資料探勘應用類(續) .....	25
表 3.1 資料探勘常用的技術 .....	42
表 4.1 變數相關分析表 .....	57
表 4.2 88、89 年選定路口之交會道路 .....	62
表 4.2 88、89 年選定路口之交會道路 ( 續 ) .....	63
表 4.3 本研究整理之路口 .....	64
表 4.4 88、89 年路口之交通事故次數 .....	66
表 4.5 92、93 年之交通事故次數 .....	67
表 4.6 相異時間路口總事故次數彙整表 .....	68
表 4.6 相異時間路口總事故次數彙整表(續) .....	69
表 4.7 台中港路與文心路路口之基本調查資料 .....	70
表 4.8 中港路與文心路路口各行向組合之交通事故實際次數 ..	71
表 4.9 忠明南路與南屯路路口之基本調查資 .....	72
表 4.10 忠明南路與南屯路路口各行向組合之交通事故實際次數	73
表 4.11 行向組合交通事故次數統計表 .....	74
表 4.12 選取路口肇事時段統計表 .....	75
表 5.1 變數資料之基本統計 .....	77
表 5.2 變數資料之基本統計 .....	79
表 5.3 負二項迴歸模式變數彈性值 .....	84
表 5.4 網路相關設定值 .....	86

表 5.5 相異演算法及隱藏層節點數之訓練驗證結果 .....	87
表 5.6 相異隱藏層節點數之訓練驗證結果 .....	87
表 5.7 類神經網路輸入變數之敏感度平均值 .....	91
表 5.8 敏感度值先取絕對值後再平均之結果 .....	93
表 5.9 類神經網路與負二項迴歸模式之比較 .....	94
表 5.10 敏感度與彈性分析之比較 .....	94
表 6.1 國內外辨識危險對象之主要方法論彙整表 .....	97
表 6.1 國內外辨識危險對象之主要方法論彙整表(續) .....	98
表 6.2 決策樹輸入變數代號 .....	101
表 6.3 Gini決策樹類別與終端節點表 .....	103
表 6.4 Twoing決策樹類別與終端節點表 .....	104
表 6.5 Gini決策樹預測正確率表 .....	105
表 6.6 Twoing決策樹預測正確率表 .....	105
表 6.7 Gini與Twoing變數重要度 .....	106
表 6.8 Gini決策樹判斷路口安全危險程度法則 .....	108
表 6.8 Gini決策樹判斷路口安全危險程度法則(續) .....	109
表 6.8 Gini決策樹判斷路口安全危險程度法則(續) .....	110
表 6.8 Gini決策樹判斷路口安全危險程度法則(續) .....	111
表 6.8 Gini決策樹判斷路口安全危險程度法則(續) .....	112
表 6.8 Gini決策樹判斷路口安全危險程度法則(續) .....	113
表 6.8 Gini決策樹判斷路口安全危險程度法則(續) .....	114
表 6.9 五權西路與環中路路口之基本調查資料 .....	116
表 6.10 五權西路與環中路路口各行向組合之交通事故實際次數 .....	116
表 6.11 輸入變數與決策樹輸出結果資料對照表(行向一與行向二) .....	117
表 6.12 輸入變數與決策樹輸出結果資料對照表(行向一與行向四) .....	118
表 6.13 輸入變數與決策樹輸出結果資料對照表(行向一與行向三) .....	119
表 6.14 輸入變數與決策樹輸出結果資料對照表(行向二與行向四) .....	119

表 6.15 輸入變數與決策樹輸出結果資料對照表(行向三與行向三)	
.....	120
表 6.16 輸入變數與決策樹輸出結果資料對照表(行向二與行向二)	
.....	120



# 第一章 緒論

## 1.1 研究背景與動機

近年來台灣地區汽、機車持有數逐年增加，根據交通部統計處於民國 89 年至 93 年近五年統計之資料，機動車輛每年平均成長率為 3.68%；道路里程成長率增為 1.47%(如表 1.1)。雖然道路里程數逐年增加，但其速率卻不及車輛成長率迅速，因此造成道路擁擠程度加劇，並增加行車之困難與危險。同時從民國 89 年至 93 年，短短五年間，台灣地區的事故件數增加 84,269 件，年成長率 14.14%，受傷人數更遽增 112,213 人，年成長率 14.59%(如表 1.2)。顯示交通事故對國民生命財產之危害，已到不容忽視之程度。

表 1.1 台灣地區機動車輛與道路里程統計

年度	機動車輛數 (萬輛)	道路總里程 (公里)	車道公里 (公里)	市區道路里程 (公里)
89 年	1,702	35,931	20,375	15,556
90 年	1,746	36,698	20,654	16,044
91 年	1,790	36,978	20,816	16,162
92 年	1,850	37,642	20,947	16,695
93 年	1,918	38,197	20,994	17,203
年平均 成長率	3.68%	1.47%	0.22%	3.04%

資料來源：交通部網頁 ( <http://www.motc.gov.tw/> ) 及本研究整理。

表 1.2 台灣地區歷年道路交通事故統計

年度	交通事故件數	肇事率 (件/萬輛)	死亡人數	受傷人數
89 年	52,952	31.11	3,388	66,895
90 年	64,264	36.81	3,344	80,612
91 年	86,259	48.19	2,861	109,594
92 年	120,223	64.99	2,718	156,303
93 年	137,221	71.54	2,634	179,108
年平均成 長率	14.14%	10.09%	-3.09%	14.59%

資料來源：交通部網頁 ( <http://www.motc.gov.tw/> ) 及本研究整理。

雖然政府努力解決交通問題，提出都會捷運化、輕軌化等方案，並極力鼓吹國民搭乘大眾運具，然而卻囿於經費所限，無法全面推廣，又興建捷運或輕軌，除了成本昂貴外，興建更加曠時費日，無法立即紓解交通擁擠與降低交通肇事率，是故要解決交通問題，必須先從交通三要素－「人」、「車」、「路」三方面著手，輔以交通三 E－教育(Education)、工程(Engineering)、執法(Enforcement)之手段；同時發展大眾運輸系統，如此雙管齊下，方能保障民眾生命財產之安全，同時能降低交通事故之肇事率。

而在人方面，係指駕駛人或行人的身心狀態，如：熬夜開車、判斷錯誤及生理上或心理上的限制等；車方面，則指車輛本身的性能，如：車輛的尺寸、穩定度、保養程度及安全設備等；至於路方面，為道路之幾何設計與交通號誌，如車道數、分隔島、道路寬度及速限或警告標誌等。

從過去研究中，Gupta 與 Mann【32】指出交通事故之發生雖係由人、車、路及環境四大因素共同作用所影響，且道路因素與交通事故之間有統計關係之存在。結果發現，車道個數、路肩寬度及交通量之增加，均使得肇事率明顯增加；Andrey 與 Yagar【36】僅考量天氣因素而不考慮下雨後所造成路面濕滑程度之情形下，發現平日事故發生機率為一定正常之機率值，一旦下雨後，同一地點發生事故的機率將比未下雨時高出 70%；Ogen, K. W.【39】提出道路的路肩有無鋪面及鋪面的寬度將對肇事有頗大之影響；Zegeer【52】分析二車道公路肇事與交通特性及公路幾何間之關係，結果發現影響肇事之重要變數有平均每日交通量、彎曲路段百分比、道路寬度、交叉路口數、溝渠數及地形為丘陵或平地。

由上述可知，影響交通事故發生之因素繁多，諸如：人為疏失、天候因素、車輛因素與道路環境等等，交通事故的形成受到上述各因素相互影響，如圖 1.1 所示。針對影響肇事因素而言，人、車兩要素，政府除了加強教育民眾正確的交通觀念，並嚴格執法外，實無太多改善手段；路方面則不同，政府可藉由重新規劃道路、路口槽化、時向

改變、調撥車道等方式來降低肇事率。故本研究僅將焦點放在路方面，藉由工程技術來針對道路作改善，期有效降低肇事率與嚴重程度，其次，亦能提供日後道路設計改良規劃之參考，評估交通管理措施之可行性。

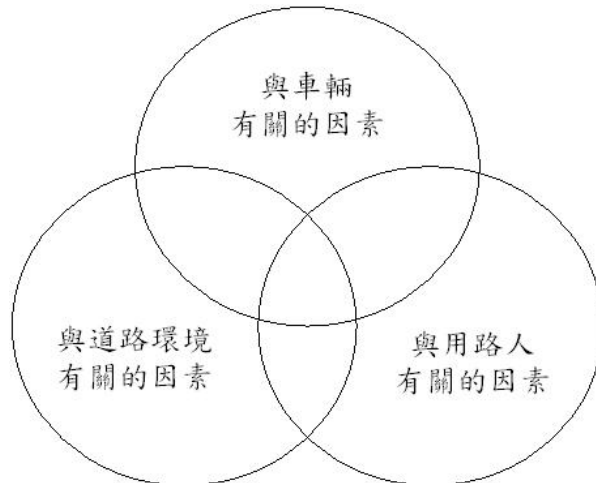


圖 1.1 人車路因素交集圖

## 1.2 研究目的

過去各種事故相關研究中，發現大部分皆透過各種不同技術來建構事故危機與否之預警模式。事實上，截至目前為止，針對此依研究議題在學術上已有許多分類之工具被發展出來，包括傳統統計方法、類神經網路(Artificial Neural Networks, ANN)及資料探勘(Data Mining, DM)等。而從過去研究發現，傳統統計方法具有以下幾點困境：

1. 選擇獨立變數納入模式之中較為困難，需依據過去文獻之分析結果。
2. 統計模式常以平均數與變異數相等為基本假設，而實際上，事故次數為計數值資料，兩母體資料間平均數難以與變異數相等。
3. 交通事件發生的時間、地點是隨機的，具有偶然性和不可預測性，而事故頻率發生較少的路口，若以行向組合區別，則多存著大量為零之次數。

故本研究則嘗試以不同的研究方法分析，雖然統計模式有其缺失存在，但仍可透過與統計模式分析結果進行比較，以評估此方法之應用可行性。相對於傳統統計方法之應用，類神經網路及資料探勘兼具備處理大量資料之特性。類神經網路沒有傳統統計方法建構模式時需要滿足許多假設條件之要求，且具有建構非線性模式之優越能力。而分類與迴歸樹(Classification and Regression Trees, CART)是另一個較新穎且較少被應用之技術與方法，利用二元分割過程來分析龐大的資料集，並透過遞迴(Recursive)的程序，產生一組劃分好的規則(Rule)，可透過此規則對樣本進行預測之工作。

從相關研究得知，交通事故已非將人為因素視為交通事故發生為唯一被考量的因素，且與其道路環境及交通工程設施因素彼此間相互影響且錯綜複雜；事故與幾何、號誌及交通狀況因素間存在高度非線性關係，無法使用線性迴歸求得相互間之合理關係。因此，本論文嘗試以類神經網路及資料探勘技術為主與統計分析模式為輔之方法，以台中市交通事故為例，利用楊憲國君於民國 88、89 年之肇事資料及民國 92、93 年肇事資料，分析其肇事資料庫和交通環境資料庫，探討影響交通事故之潛在因素。故本論文的研究目的為：

1. 分析肇事路口之主要肇事原因。
2. 建構統計、類神經網路對於路口交通事故之預測模式，並探討道路環境與交通工程設施對路口交通事故之重要潛在因素。
3. 對所得模式進行分析，期瞭解及比較該等研究方法對分析交通事故問題之適用性。
4. 考量現行作業對於肇事路口主要肇因之判定並無一套完整之作業規範，利用分類與迴歸樹產生一組劃分好的規則判定，以降低易肇事地點之危險程度。

### 1.3 研究範圍

相較路口及路段發生之交通事故位置而言，路口為最常發生之事故地點，而路段之事故地點往往過於分散，欲界定路段事故發生之範



圍實屬不易且難以區分道路環境與交通工程設施特徵。基於此考量下，本研究針對台中市路口發生之事故資料為研究範圍。而一般發生於交岔路口內或交岔路口附近之交通事故，係指兩條以上道路平面交岔重疊部份及從該部份向外延伸 10 公尺以內之範圍。本研究所定義交岔路口範圍，即參考上述係指兩條以上道路平面交岔重疊部分及從該部份向外延伸 10 公尺以內之範圍。以四岔路口為例，形成一個涵蓋五個區域的交岔路口平面圖，如圖 1.2 所示。本研究即對此一研究範圍內之交通事故進行分析研究，以期針對市區路口幾何特性，找出交通事故發生之影響因素。

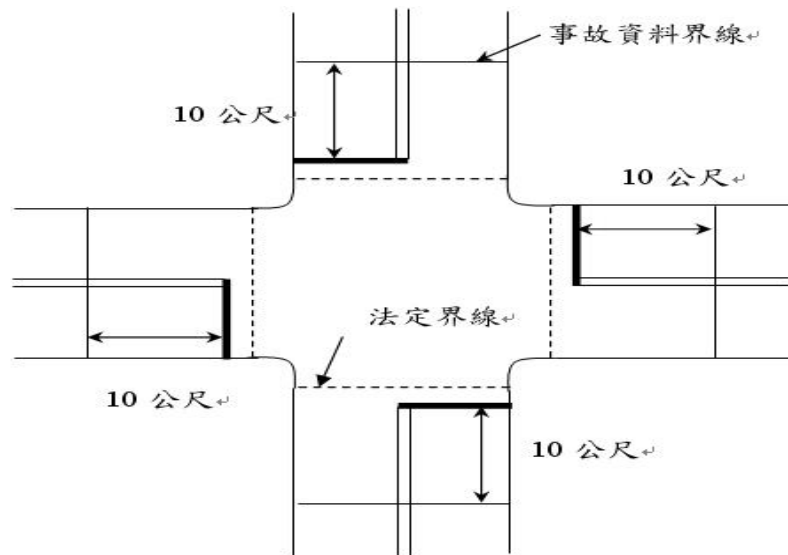


圖 1.2 路口範圍平面圖

資料來源：【楊憲國，2002】

## 1.4 研究內容與流程

本研究之研究內容如下：

### 1. 界定研究動機與目的

因交通事故之影響因素彼此間錯綜複雜及影響程度不易界定，非能以線性迴歸求得相互間之合理關係，而類神經網路及資料探勘皆具處理複雜且非線性關係之優點，針對路口交通事故及其幾何、號誌及交通現況因素進行分析研究，建構一個有效之預測模式，並找出交通事故發生之影響重要因素。

## 2. 文獻回顧

### (1).傳統迴歸

彙整國內外有關以傳統統計方法分析交通事故之研究方法與結果，比較其不同分析方法間之差異性與所帶來之效果，並整合相關文獻中影響肇事鑑定之重要因子，藉此整理其相關變數以瞭解當前實際狀況，並決定本研究之交通事故分析之方法。

### (2).類神經網路

彙整國內外有關類神經網路(Artificial Neural Networks, ANN)相關應用文獻，以學習其網路架構形式與學習法則，藉此瞭解文獻中的類神經網路之功用與特性，作為本研究模式之架構方式與預測之準備。

### (3).資料探勘

彙整國內外有關資料探勘(Data Mining, DM)相關應用文獻，以學習其分類之演算法，透過文獻中資料探勘之功用與特性，作為本研究所採用之模式演算法應用於預測之準備。

## 3. 統計分析評選重要之變數及蒐集事故資料

本研究主擬以楊憲國君之台中市 62 個路口事故數據，選取 29 種變數加以探討事故發生之次數。事實上，決策者僅瞭解事故發生所造成之影響程度何者為重即可，由於因素彼此間錯綜複雜及影響程度不易界定，故本研究先行統計分析之相關性分析，解決變數共線性之問題，以增加其可信度；再挑選近年內台中市其他路口之事故，依初步篩選重要因素為準則，整理成模式所需資料型式。

#### 4. 模式建構

##### (1).統計模式

經由整理交通事故資料與評選出重要變數後，利用 Limdep 7.0 先以檢定本模式適宜卜瓦松迴歸或負二項迴歸，並建構事故預測模式，並針對其變數之彈性，檢驗其變數間之顯著關係，求得較精準之預測結果。

##### (2).類神經網路模式

經由整理交通事故資料與評選出重要變數後，利用 Matlab 6.5 編撰以試誤法方式來建構倒傳遞類神經網路模式之架構，擇選較佳模式後，再進行敏感度分析，根據分析結果探討變數對事故之影響。

#### 5. 分類與迴歸樹

交通事故之發生具有偶然性與不可預測性，加上影響因素眾多且彼此間相互影響，即使同一因子對於不同路口之行向組合影響亦有所不同。為瞭解因素在各行向組合中如何影響路口之危險程度，因此經由統計模式、類神經網路分析評選出重要變數，透過適宜的道路潛在危險性評估指標，利用 C 4.5 探討資料探勘技術之分類與迴歸樹分析影響交通事故嚴重程度之因素及其影響程度，產生一組劃分好的規則，以判定肇事地點之危險程度，並挖掘各變數在路口危險程度中所扮演之角色。

#### 6. 結論與建議

經過瞭解影響事故結果之重要因素後，再綜合上述之研究成果提出具體結論與建議，以期能給予決策者規劃路口之優先考量準則。

本研究之研究流程如下圖 1.3 所示。

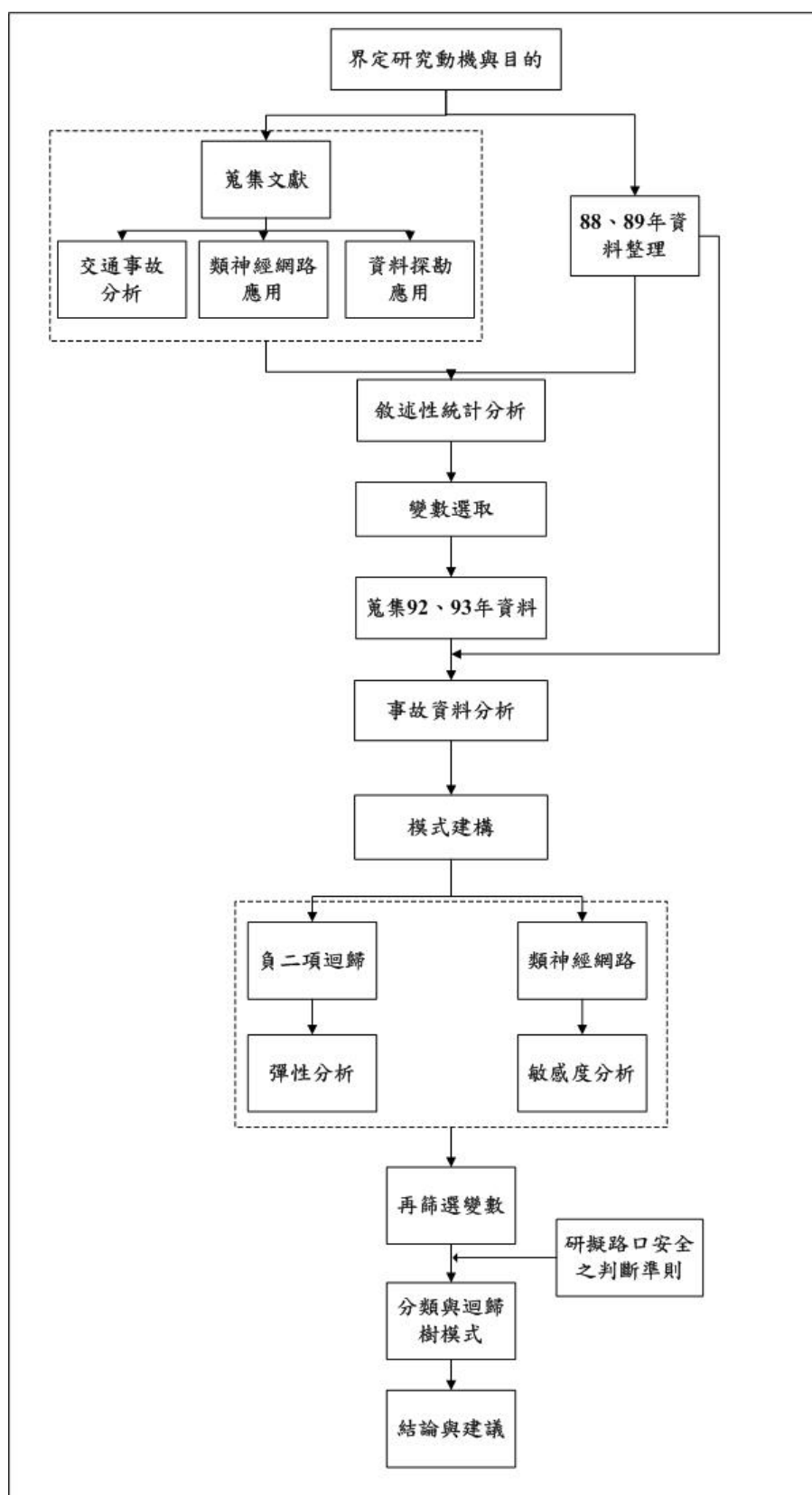


圖 1.3 研究流程圖

## 第二章 文獻回顧

本章文獻分為三節，分別探討國內外以傳統模式分析交通事故之嚴重程度，回顧類神經網路與資料探勘之應用領域，瞭解可行之研究方法以作為本論文研究方法之參考。

### 2.1 傳統模式分析交通事故之嚴重程度

過去關於交通事故方面，評估方法以迴歸統計模式居多。其中，最常見之統計迴歸模式可分為線性迴歸(Linear Regression)、卜瓦松迴歸(Poisson Regression)、負二項迴歸(Negative Binomial Regression)及羅吉斯特迴歸(Logistic Regression)等。早期係採用線性迴歸模式，近年來有鑑於交通事故之稀少及非負數特性與常態分配假設相違，而以卜瓦松迴歸及負二項迴歸較常採用，鑑於資料常以類別資料呈現，羅吉斯特迴歸亦陸續被使用，以下將敘述各類模式之相關文獻。

Moses et al. (1994)探討貨車行特性與肇事之關係，考慮到卜瓦松迴歸模式之平均數等於變異數假設之不合理性，經檢定發現平均數確實不等於變異數後，改採以負二項迴歸模式建構模式。因變數為貨車肇事率，自變數為貨車行規模、年資、載貨之貨物特性、車隊行駛里程及是否服從安全管制等。發現除支持之前研究所獲致結論外，尚發現肇事與貨車行規模無關，卻與其經營年資有關；載運受託貨物、載運危險物質都易導致較高肇事率。

Shanker et al. (1995)認為卜瓦松迴歸模式適合於肇事發生之分析，但亦須符合平均數與變異數相等之假設。然而資料的過度偏離(變異數大於平均數)或較少偏離(變異數小於平均數)將導致參數估計值之偏誤，而負二項迴歸模式則可以克服此資料的偏誤。以負二項迴歸模式建構交通肇事頻率與道路幾何、氣候及其他季節性因素間之關係，結果顯示道路幾何設計與氣候之交互關係會影響交通肇事之頻率。

Parker et al. (1995a)透過多元線性迴歸(Multiple Linear Regression)方法來預測駕駛者之駕駛行為對於事故發生之影響，以駕駛者行為問卷(Driver Behavior Questionnaire, DBQ)、駕駛者風格問卷(Driver Style

Questionnaire, DSQ)及駕駛者決策問卷(Decision Making Questionnaire, DMQ)等三種問卷來蒐集英國地區駕駛者之違規行為、錯誤(Negligences)、失誤(Lapses)、行車速度及行駛時駕駛者的決策方式等行為資料，以探討駕駛者的違規行為、開快車及缺乏足夠的判斷決策能力等特性是否會增加不同交通事故型態的發生。結果顯示無論主動或被動違規行為均造成事故發生有關；而開快車及缺乏決策者與主動事故發生有關；開快車與直角對撞事故發生有關等。

Parker et al. (1995b)探討駕駛者違規與事故之關聯性，亦採用駕駛者行為問卷(DBQ)來蒐集英國地區駕駛人的駕駛行為及事故發生次數等資料，採多元線性迴歸方法來預測駕駛者之肇事率，探討駕駛者的失誤、錯誤及違規行為對於事故發生的影響。結果顯示違規行為會增加肇事率，而失誤及錯誤則無顯著之影響肇事率。

Shankar et al. (1997)探討路段事故稀少之問題，認為發生頻率記錄為零時，不必然代表該路段較安全，其可能較安全而未發生事故或曾發生事故卻並未記錄，亦可能調查時間並未發生事故。以假設此發生機率值為常態或羅吉斯特分配來分析事故頻率之資料，利用ZIP(Zero-Inflated Poisson)與ZINB(Zero-Inflated Negative Binomial)迴歸模式將事故次數為零的情況設定一機率值，結果顯示在主要幹道之肇事頻率上，以負二項迴歸較適用，主要的影響因素有年平均日交通量、路段長度、道路坡度、車道個數及路段速限等；而在次要幹道之肇事頻率上，以ZINB迴歸模式較適用，主要影響因素有年平均日交通量、路肩寬度、曲率半徑、道路坡度及直線路段等；而在連絡幹道之肇事頻率上，則以ZIP迴歸模式較適用，主要影響因素有年平均日交通量、路段長度及曲率半徑等。

林郁志君(1998)選定高肇事率路段及路口蒐集相關因子，並利用一般線性模式中的卜瓦松迴歸模式，嘗試找出與肇事因果相關之因素，發現交通事故之發生，主要與駕駛人行為特性、車輛特性、道路幾何特性及環境因素等之相關性較高。

Abdel-Aty et al. (1999)針對駕駛者與道路設計於交通事故對於事

故發生與涉入進行探討。除調查在高速公路上影響事故事件的因素外，進一步針對駕駛者之性別及年齡進行討論。利用多元線性迴歸(Multiple Linear Regression)、卜瓦松迴歸(Poisson Regression)及負二項迴歸(Negative Binomial Regression)進行分析，變數選取有平曲線、路肩寬度、中央分隔帶寬度、交通量及速限，其結果發現 AADT 為共同主要的因素，僅次為道路寬度及車道個數；女性駕駛者肇事率較男性高；年輕與老年駕駛者肇事機率較中年駕駛者高。

Al-Ghamde (2002)探討影響交通事故嚴重之主因，以交通警察單位所記載之 560 筆交通事故資料的受傷者為樣本，透過對照查詢醫療報告來判斷受傷者之嚴重程度，將嚴重程度分為受傷與死亡兩類，利用羅吉斯特迴歸來分析事故地點及肇事主因與事故嚴重程度之關聯性。結果顯示事故地點及肇事原因與事故嚴重程度具有顯著相關，而非號誌化路口之事故嚴重度顯著高於號誌化路口；造成肇事歸咎於駕駛者闖紅燈、逆向行駛單行道，其中又以逆向行駛單行道之事故死亡勝算比遠高於其他肇事原因，表示逆向行駛單行道的駕駛人事故死亡的機率高於其他肇事原因的駕駛者。

Al-Ghamdi (2003)利用敘述統計學、成對樣本比較、卡方獨立性檢定及曝光量不易獲取下以條件機率，分析沙烏地阿拉伯首都利雅得於交叉路口發生交通事故傷害，瞭解在交叉路口或是非交叉路口對於交通事故之問題，以能避免事故發生，進而供改善交通問題之對策。因此資料中總共有 1,774 件意外事故，時間為 1996-1998 間，事故因素有事故型態，碰撞型態，原因和時間，駕駛者速度和其他重要的行為因素。結果顯示 50%嚴重事故都涉及行人，而主要發生在非交叉路口上。不適當的駕駛行為為背負事故之重則，如超速、無讓路、闖紅燈比其他所有事故因素的機率皆高於一半以上。

## 2.2 類神經網路

類神經網路應用於醫學、資訊、工程及金融等其它領域之研究相當多，其中應用於醫學與金融業方面與個人特性因素之相關性最高。近年來，國內交通運輸領域也逐漸重視類神經網路之功能，且有相當

之研究成果。在交通領域上多應用類神經網路於分類、歸類、尋求最佳解、預測功能等。本研究將類神經網路之預測特性於交通運輸課題之應用及其他領域整理如下：

### 2.2.1 交通運輸課題之應用

Hashemi et al. (1995)基於 Hebb's(1949)學習法則，採倒傳遞網路來預測密西西比河下游船舶意外事故與環境因素間之影響，並與多元判別分析、羅吉斯特迴歸相比，以瞭解船舶肇事預測與績效。分別建構以船與船之間碰撞(Collision)、船撞到固定物(Ramming)及船身底部碰撞事故(Grounding)三種事故之預測準確度，結果顯示神經網路預測率為(67%，77%，90%)，區別分析為(25%，62%，65%)，及羅吉斯特迴歸為(33%，62%，65%)。因此類神經網路提供了較佳之預測績效，並優於其他兩模式之預測結果。

Ivan (1997)探討芝加哥郊區幹道研究道路發生事件所造成交通擁擠之問題，以迴圈式固定偵測系統(FDA)與探測車(PVA)兩種偵測系統配合 two input network、output memory network、adjacent link network、full network 四種倒傳遞網路方法。驗證發現 full network、output memory network 內皆含 memory unit 可進一步篩檢預測值而得較佳之偵測率 92.59%、86.11%。經此四個 NN 模式道路事件偵測資料可得到很好偵測率及 0.01% 較低的錯誤預警頻率(False Alarm Rate, FAR)，且 NN 模式於資料轉換上能較其他方法能有效降低誤判率。

Subba Rao et al. (1998)以類神經網路分析運具選擇問題之適用性，以印度孟買政府對於居民對到達鐵路車站搭乘火車之前運輸需求(包括步行、公車、機車、汽車及計程車等運具)調查結果為樣本資料，利用多重羅吉(Multinomial Logit)及倒傳遞網路模式分別建構運輸需求預估模式，並比較傳統選擇模式與類神經網路的好壞，其結果顯示類神經網路模式有較佳的預估效果。

Mussone et al. (1999)為評估交叉路口發生事故之影響因素，考量因素有道路因素、能見度、天候因素、駕駛者及運具。所使用的資料為米蘭的事故資料庫，時間為 1992-1995 年之資料，共 46,000 件事故



發生在交叉路口，其資料庫組成之特性有環境因素、碰撞因素和道路幾何設計因素。而此研究之方法是用類神經網路，輸入層為環境因素、碰撞因素和道路幾何設計因素，輸出層則為事故次數。結果顯示有號誌化路口中，夜間發生碰撞事故機率較高；非號誌化路口則以夜間行人發生事故機率較高。

周雍傑君(2000)探討市區道路肇事之嚴重程度，以駕駛人資料、車輛資料、道路幾何及道路設施為基本資料，透過倒傳遞網路，將路口與路段分開建構模式，發現類神經網路模式在路口、路段的肇事嚴重程度之正確率可達八成以上，其敏感度分析顯示在路口上，當天候不佳且潮濕路面及路面有障礙物時，肇事死亡比例則會提高；而在路段上，當天候不佳及潮濕路面時，肇事死亡的比例則會提高。

李穎(2002)利用類神經網路於國道客運班車旅行時間預測模式上，其針對高速公路上客運之 GPS 資料、道路上面的車輛偵測器資料來判讀目前的車速及流量等輸入因素，觀察高速公路之運行情況及評估分析旅行時間。並以不同的輸入因素組合搭配不同的轉換函數，以求取較佳的預測模式，其預測的結果誤差在 11.3%~25.4%間，處於優良之預測水準。

楊憲國君(2002)探討交通事故發生次數預測，以道路幾何、號誌設施及交通流量等因素為輸入變數、交通事故發生次數為輸出變數，透過類神經網路與負二項迴歸方法來架構模式，比較類神經網路與負二項迴歸模式之預測能力，其結果顯示類神經網路之預測能力較佳，其最大誤判率為 16.4%。

吳志文(2002)探討船舶事故之嚴重程度，以民國 75-89 年間之「高雄港歷年重大海事評議案例」及高雄港船舶事故案件為資料，以季節、時間、噸位、位置、船舶類型、引水人、肇事因素、風速、風向、浪高為輸入變數，分別以單元及多元神經元編碼建構船舶事故嚴重程度模式。結果發現以多元神經元編碼之船舶事故型態，平均總正確率為 85.78%(碰撞 96.55%、其他(擱淺，沉沒)為 75.00%)，以單元神經元編碼之船舶事故型態，平均總正確率為 72.41%(碰撞 94.83%、其他(擱

淺，沉沒)為 50.00%)，顯示類神經網路應用於預測船舶事故亦有良好的績效。

莊智仁(2003)針對民國 90 年發生於台北市的交通事故及台北地區近五年駕駛人之違規記錄資料進行抽樣，利用 1:1 配對資料之羅吉斯特迴歸模式與類神經網路模式之方法，分析各種不同違規行為對交通事故發生之影響並預測違規駕駛人潛在發生事故之機率。發現許多違規因素會顯著的影響肇事發生，其中具有闖單行道、酒後駕車、車輛設備未依規定及未禮讓直行車或行人者等違規行為的駕駛人有較高潛在之肇事機率；而車輛未定期檢驗、未戴安全帽及未戴安全帶等違規行為則未明顯與事故發生相關。綜合比較類神經網路與羅吉斯特迴歸分析之結果，類神經網路之預測正確率略高於羅吉斯特迴歸模式。

黃昶斌(2004)建構一個類神經網路模式，用以預測出某路口、路段在發生肇事後，當事人之傷亡程度，測試結果顯示，路口肇事嚴重程度預測模式對於一般受傷案件的預測正確率為 95%，重傷、死亡案件的預測正確率為 42%，而路段肇事嚴重程度預測模式對於一般受傷案件的預測正確率為 92%，重傷、死亡案件的預測正確率為 45%，顯見類神經網路方法所建構的預測肇事嚴重程度模式有相當良好的績效。另外將模式實際應用於易肇事道路，所預判出會造成嚴重傷亡肇事案件的違規類型，可及早對管理者提出警訊，並研擬適當的行政進而加以因應。

劉秀鳳(2005)認為地震後易造成邊坡土質鬆動，以至於颱風或豪雨過後，總帶來許多邊坡崩塌災害。研究以台 14 線 31k+800~75k+200 路段 13 處易發生崩塌之邊坡為調查地點，利用類神經網路研究各影響因子對其崩塌潛勢之影響。研究結果顯示，經由網路之訓練與測試，證明類神經網路對 13 處易崩塌邊坡 462 筆全部案例崩壞與否之正確率達 90%。

### 2.2.2 其他

王信勝(2001)為建立銀行對於顧客的信用評分制度，以現行銀行內部現有的客戶信用評分表資料，配合客戶的繳款記錄，建構客戶的

信用評資料與客戶貸款狀態的資料表，利用羅吉斯特迴歸、倒傳遞類神經網路及層級分析法結合類神經網路的方法來建構是否是好貸款的模式。其結果顯示三種模式中以層級分析法結合類神經網路方法的正確率最高。

陳啟元(2003)為建立醫院門診點數預測模式，而讓健保局合理分配醫療資源之參考。以北區健保局所屬之各級醫院自民國 89-91 年間之門診點數清單申報資料為樣本。自變數與因變數的關係上，以某期影響因子(不同性別人數、不同年齡分組人數、醫師人數等)與該期之下一期之服務點數配對，以建立各層級及各區域級以上醫院之點數預測模式。研究結果，使用類神經網路建立之預測模式，與利用迴歸分析所建立之預測模式相較，預測值落在實際值 10%區間內的個數百分則比較高，在醫學中心層級方面，佔驗證資料總數的 74.36%(迴歸分析為 51.29%)，顯示類神經網路建立之預測模式優於迴歸分析所建立之預測模式。

謝森松(2004)針對於醫院發生跌倒之意外事件，導致病人病情惡化、產生併發症及延長其住院日數，嚴重耗用醫療成本。因此藉由架構類神經網路模式，進一步地準確預測住院跌倒之發生，確保病人安全及降低醫療費用。以台北市某醫院中心民國 90-92 年住院跌倒之電子病歷資料，將案例中的 2,796 筆當作訓練組，350 筆當作測試組，結果顯示所採之類神經網路模式所得之敏感性與特異性則分別為 87.19%及 87.64%，顯示電子病歷在預測住院病人跌倒之發生可以作為良好的資料來源。

## 2.3 資料探勘部分

『資料探勘技術是近年來隨著人工智慧與資料庫技術的發展，而出現的一門新興技術。』【5】其應用領域之廣泛，更包含許多不同之分析方法。最常被用於商業行為上，如保險業、銀行業與零售業，近年來也被用於其他領域，如醫療業、製造業、生化與農業上，而交通安全領域應用卻十分罕見，因此本研究將嘗試將資料探勘技術應用於交通事故因素分析上。應用整理如下：

### 2.3.1 交通運輸課題之應用

Kuhnert et al. (2000)主要研究車輛碰撞造成嚴重損傷之嚴重性因素，針對非參數模式分類與迴歸樹(Classification and Regression Trees, CART)、MARS(Multivariate Adaptive Regression Splines)和羅吉斯特迴歸模式之三種研究方法分析，蒐集於澳洲布里斯班之汽車運具意外事故，事故樣本是利用與病人面談得知之資料，時間為 1997-1998 年，利用問卷方式，針對幾年的駕駛經驗、駕駛安全預防、駕駛者侵略行為(Driver Aggression)、動動尋找行為(Thrill Seeking Behaviour)和一些基本社經資料。其次，再經由住院六個月後之後續問卷進行分析，而此問卷則是包含駕駛經驗，駕駛者意見和行為，樣本資料則為 2,000 位受害者，年齡為超過 17 歲的人。羅吉斯特迴歸在分析時以 663 個訓練樣本及 55 個測試樣本，CART 和 MARS 有 689 個訓練樣本和 59 個測試樣本分析。結果發現羅吉斯特迴歸總正確率小於 CART 和 MARS，以 MARS 而言，可以得到變數之間的交互項；CART，可以提供主要變數分割的訊息，而羅吉斯特迴歸將變數分割的訊息分析，此研究再將此三個模式做結合，同時進行分析時，則可以有更好的預測能力。

Karlaftis et al. (2002)探討道路幾何設計與交通肇事發生之關係，透過非參數統計方法(Non-Parametric Statistical Methodology)－階層樹狀迴歸(Hierarchical Tree-Based Regression, HTBR)來分析與預測交通肇事的發生。資料為印第安那州運輸部，資料範圍 1991-1995 年所發生在郊區道路的交通事故，而道路資料方面，包括交通量資料、鋪面、車道寬、號誌控制情形等。研究結果發現，年每日平均交通量(AADT)為一個重要之變數，因此樹狀結構中以 AADT 判斷法則為樹之根部，而且出現的節點次數最多，所以 AADT 判斷交通事故的發生為必要條件，次要的預測變數為車道寬度、煞車係數等。由於 HTBR 分析方式能快速找到肇事因素，對於交通工程師在設計道路上更有幫助；另外，在交通安全管理上也可以對已知的路段特性下去預測其發生肇事之次數，採取相關之因應措施。

陳文杰君(2004)探究高速公路交通意外與發生原因之重要關係，

透過分類與迴歸樹(Classification and Regression Trees)與負二項迴歸(Negative Binomial Regression)進行分析,發現在客車交通量大於4,677輛之路段,會有較高的機率發生2次事故,在2,096至4,677輛之路段則發生1次事故之機率較高,小於2,096輛之路段則不會發生事故之機率較高,此結果負二項迴歸模式並無法分析出。預測率方面,決策樹的正確率有58%而負二項迴歸模式有53%,兩者差異不大,證實資料探勘方法為可行之分析工具。

王秀雯君(2004)針對台北市發生交通事故的受傷者為觀測者進行分析,時間為2001年間之資料分析。嘗試利用資料探勘技術為主與統計分析模式的方法為輔以進行分析,以資料探勘技術中分類與迴歸樹及多元羅吉斯特模式分析造成交通事故嚴重程度之因素,最後,比較兩種模式對於影響事故嚴重程度之變數及預測分類正確率。結果顯示整體預測能力以決策樹優於多元羅吉斯特模式,而事故嚴重程度之變數,多元羅吉斯特與決策樹大致相同。顯示資料探勘可以適用於運輸安全領域。

### 2.3.2 其他

Marshall (2001)嘗試找出影響心肌衰竭、梗塞及糖尿病病患等高風險因素之使用模式,以建立診斷結果和預測症狀規則機制,用以評估在臨床流行病學分析上之高風險群組。過去文獻採用「自動交互偵察」(Automatic Interaction Detection, AID)在分析資料可以揭露同性質性母體之子群組之間的關係,但可能會辨別出官能的雜訊(Make Send Out of Noise)或是潛在偽造的結果,而毫無影響的因素會被決定出來,不過分類與迴歸樹的方法可以處理問題的過適性,且可以採用成本特異錯綜性(Cost-Complexity)找出最為接近適化之資料。因此,此研究則介紹以決策樹方法以建構好的迴歸模式,建立診斷結果和預測症狀規則機制。

彭慧雯君(2001)為建立銀行之信用卡發卡原則,利用資料探勘(Data Mining)的方法從銀行的信用卡用戶資料中尋找重要的變數資訊,配合應用區別分析法(Discrimination Analysis)與倒傳遞類神經網

路建構信用卡客戶的分類模式，做為銀行發卡審核、計算發卡風險及收益之用。結果顯示無論區別分析與類神經網路之結果均不佳，透過增加變數個數、改變樣本中族群比例與樣本數目可提高類神經網路的分類正確率。

Sohn et al. (2003)探討資料融合、總體和群聚分類於韓國道路交通事故準確度之研究，以道路寬度、車型、事故類型(A1、A2)、事故前車速、激烈駕駛及保護設備作為輸入變數，分別以類神經網路(Artificial Neural Networks)、決策樹(Decision Tree)、丹波斯特-雪佛爾算法(Dumpster-Shafer)、貝氏估計法(Bayesian)及羅吉斯特模式(Logistic)進行比較分析。結果顯示單一分類比較時，貝氏估計法結果最具有代表性。比較發現，丹波斯特-雪佛爾算法能改善個別的分類性能，如神經網路和決策樹，亦融合貝氏估計法和羅吉斯特。

Bevilacqua et al. (2003)分析唧筒(Pump)的錯誤率增加之因素及唧筒之操作錯誤而影響錯誤率增加之因素，由於高信賴水準以及高品質的唧筒設備對煉油廠為必要之條件，通常預期造成失誤率的電子元件之條件有溫度、振動水準、溼度等都是潛在失誤因素，因此在研究方法是分類與迴歸樹方法分析，目的在於辨別主要影響 143 個不同的離心唧筒(Centrifugal Pumps)的操作條件中會影響錯誤率增加之因素，資料範圍是在義大利的煉油廠，所有失誤之發生共有 18 個月記錄。最後結果是所得到的解釋變數都是影響唧筒的錯誤率增加，如影響失誤率有機器設備型態為最主要因素，再者是焊接、液體、煤炭等變數都是影響操作條件之因素。

楊宗彥君(2003)探討股票價格應反映於公司之價值，藉由資料探勘中的類神經網路與決策樹技術，針對上市公司的財務比率來預測股票投資報酬。以 1994 年第一季至 2002 年第三季之電子類股作為資料，公司之各項財務比率資料為自變數，季股價報酬率為因變數，以求方法對於股票投資報酬預測之準確度。結果顯示類神經網路之準確度比決策樹要高(類神經網路 71.49%，決策樹 64.25%)，而不論類神經網路或是決策樹，皆可以透過增加樣本數或增加自變數之方式來提高準確

度，而變數之增加所帶來之效益大於增加樣本數(新增變數：類神經網路 79.64%，決策樹 71.95%；新增樣本數：類神經網路 64.25%，決策樹 57.46%)。

尤春惠君(2004)於醫療研究上，為了使醫療機構能從血液濃度監測程序(TDM)協助醫藥人員監控藥物在病患身上的作用結果，藉以調整藥物的用量與使用方式。嘗試以資料探勘技術中的 C4.5 決策樹分析法及倒傳遞類神經網路，對於泛可黴素進行 TDM 監控。由 2000-2004 觀察到的 87 筆之記錄進行分析，所牽涉之變數有性別、年齡、體重、血清尿素氮值(BUN)、肌酐值(SCR)、用藥劑量、投藥間隔等，建構可用以預測泛可黴素在病患上之作用結果之分類模式。結果顯示利用 Bagging 及 AdaBoost 兩項分類效能提升技術可提升以 C4.5 或倒傳遞類神經網路所建構之單一分類器，其中利用 C4.5+Adaboost 所建構之模式，整體正確辨識率達 79.65%，比原先專家預測正確率之 41.38% 更準確預測用量之適當性。

郭一聰君(2005)應用決策樹與類神經網路之探勘技術，根據客戶基本資料與交易活動後之應收帳款資料，組成分析變數，建立逾期預警模式，提供企業於應收帳款逾期之分析，與信用額度的檢討與設定，以降低企業呆帳的機率。研究結果發現「交易後收款天數」、「收款風險評等」、「客戶類別」、「加盟時間」、「失業率」、「前 6 月交易次數」等 6 項變數，以決策樹與類神經網路建立的逾期預警模式，確實在經銷商客戶的應收帳款管理上，有不錯之逾期預測結果。

## 2.4 文獻回顧小結

綜合上述文獻彙整如下表：

表 2.2 文獻彙整表(一)交通事故分析類

作者(年份)	研究對象	研究方法	研究結果
Moses et. al. ( 1994 )	貨車行特性 與肇事關係	卜瓦松迴歸 負二項迴歸	◆發現平均數確實不等於變異數，改採負二項迴歸。 ◆載運貨物及危險物質都易導致較高肇事率。
Shanker et al. ( 1995 )	交通肇事	負二項迴歸	◆道路幾何設計與氣候間之交互關係會影響交通肇事頻率。
Parker et al. ( 1995a )	駕駛行為對 事故之影響	多元線性迴歸	◆無論主動或被動違規行為均造成事故發生有關。 ◆而開快車及缺乏決策者與主動事故發生有關。 ◆開快車與直角對撞事故發生有關。
Parker et al. ( 1995b )	駕駛者違規 與事故之關 聯性	多元線性迴歸	◆違規行為會增加事故肇事率，而失誤及錯誤則並沒有顯著的影響事故肇事率。
Shankar et. al. ( 1997 )	路段事故	ZIP 與 ZINB 迴歸 負二項迴歸	◆主要幹道之肇事頻率上，以負二項迴歸較適用。 ◆次要幹道之肇事頻率上，以 ZINB 迴歸模式較適用。 ◆連絡幹道之肇事以 ZIP 迴歸模式較適用。



表 2.2 文獻彙整表(一)交通事故分析類(續)

作者 ( 年份 )	研究對象	研究方法	研究結果
林郁志 ( 1998 )	高肇事率路段及 路口	卜瓦松迴歸	◆駕駛人行為特性、車輛 特性、道路幾何特性及環 境因素等之相關性較高。
Abdel-Aty et. al. ( 1999 )	駕駛者與道路設 計對交通事故	多元線性迴歸 卜瓦松迴歸 負二項迴歸	◆AADT 為共同主要的 因素，僅次為道路寬度及 車道數。 ◆女性駕駛者肇事率較 男性高。 ◆年輕與老年駕駛者肇 事機率較中年駕駛者高。
AI-Ghamdi ( 2002 )	交通事故	羅吉斯特迴歸	◆非號誌化路口之事故 嚴重度顯著高於號誌化 路口。 ◆駕駛者闖紅燈、逆向行 駛單行道會增加事故死 亡機率。 ◆逆向行駛單行道之事 故死亡勝算比遠高於其 他肇事原因。
AI-aGhamdi ( 2003 )	交叉路口肇事	敘述統計學 成對樣本比較 卡方檢定	◆非交叉路口上為肇事 率較高之地點。 ◆不適當的駕駛行為比 其他所有事故因素的機 率皆高於一半以上。

表 2.3 文獻彙整表(二)類神經網路應用類

作者 ( 年份 )	研究對象	研究方法	研究結果
Hashemi et al. ( 1995 )	密西西比河 下游船舶 意外事故	倒傳遞網路 多元判別分析 羅吉斯特迴歸	◆類神經網路之預測優 於其他兩模式之預測結 果。
Subba Rao et al ( 1998 )	運具選擇問題	多重羅吉特 倒傳遞網路	◆類神經網路模式之預 估效果較佳。
Mussone et al. ( 1999 )	交叉路口 發生事故	倒傳遞網路	◆有號誌化路口中且夜 間發生碰撞事故機率較 高。 ◆非號誌化路口夜間行 人發生事故機率較高
周 雍 傑 ( 2000 )	市區道路之肇事	倒傳遞類網路	◆類神經網路模式在路 口、路段的肇事嚴重程度 的判斷可達八成以上的 正確率。 ◆天候不佳，潮濕路面及 路面有障礙物時肇事死 亡比例則會提高；路段 上，天候不佳及潮濕路面 時，其比例則會提高。
Kuhnert et al. ( 2000 )	車輛碰撞造成嚴 重損傷	分類與迴歸樹 MARS 羅吉斯特迴歸	◆羅吉斯特迴歸總正確 率小於 CART 和 MARS。 ◆以 MARS 而言，可以 得到變數之間的交互 項；CART，可以提供主 要變數分割的訊息，而羅 吉斯特迴歸將變數分割 的訊息分析。

表 2.3 文獻彙整表(二)類神經網路應用類(續)

作者 ( 年份 )	研究對象	研究方法	研究結果
王信勝 ( 2001 )	銀行對於顧客的 信用評分制度	羅吉斯特迴歸 倒傳遞類網路 層級分析法結 合類神經網路	◆層級分析法結合類神經網路之正確率高於其他兩種。
李穎 ( 2002 )	國道客運班車旅 行時間預測	倒傳遞網路	◆預測的結果誤差在 11.3%~25.4%間，處於 優良之預測水準。
楊憲國 ( 2002 )	交通事故	倒傳遞網路 負二項迴歸	◆其結果顯示類神經網 路之預測能力較佳，其最 大誤判率為 16.4%。
吳志文 ( 2002 )	船舶事故	倒傳遞網路	◆顯示類神經網路應用 於預測船舶事故亦有良 好的績效。
陳啟元 ( 2003 )	醫院門診點數，讓 健保局合理分配 醫療之資源	迴歸分析 類神經網路	◆類神經網路建立之預 測模式優於迴歸分析所 建立之預測模式。
黃昶斌 ( 2004 )	路口、路段肇事	類神經網路	◆類神經網路方法所建 構的預測肇事嚴重程度 模式有相當良好的績效。 ◆將模式實際應用於易 肇事道路，可及早對管理 者提出警訊，並研擬適當 的行政作為加以因應。

表 2.3 文獻彙整表(二)類神經網路應用類(續)

作者 ( 年份 )	研究對象	研究方法	研究結果
謝森松 ( 2004 )	醫院發生跌倒 之意外事件	類神經網路	◆類神經網路模式所得 之敏感性與特異性則分 別為 87.19%及 87.64%。
劉秀鳳 ( 2005 )	邊坡崩塌	類神經網路	◆網路之訓練與測試，證 明類神經網路對崩塌邊 坡正確率達 90%。
Marshall ( 2001 )	心肌衰竭即梗 塞、糖尿病病患 等高風險因素	分類與迴歸樹	◆此研究則介紹以二元 樹方法以建構好的迴歸 模式，建立診斷結果和預 測症狀規則機制。
彭慧雯 ( 2001 )	銀行的信用卡 發卡原則	資料探勘 區別分析 倒傳遞網路	◆透過增加變數個數、改 變樣本中族群比例與樣 本數目可提高類神經網 路的分類正確率。

表 2.4 文獻彙整表(三)資料探勘應用類

作者 ( 年份 )	研究對象	研究方法	研究結果
Sohn et al. ( 2003 )	高速公路事故	類神經網路 決策樹 丹波斯特-雪 佛爾算法 貝氏估計法 羅吉斯特	◆單一分類比較時，貝氏 估計法結果最具有代表 性。 ◆分類比較發現，丹波斯 特-雪佛爾算法能改善個 別的分類性能。
楊宗彥 ( 2003 )	股票的價格應反 映於公司的價值	類神經網路 決策樹	◆類神經網路之準確度 比決策樹要高。 ◆類神經網路或是決策 樹，皆透過增加樣本數之 方式來提高準確度。

表 2.4 文獻彙整表(三)資料探勘應用類(續)

作者 ( 年份 )	研究對象	研究方法	研究結果
楊宗彥 ( 2003 )	股票的價格應反映於公司的價值	類神經網路 決策樹	<p>◆類神經網路之準確度比決策樹要高。</p> <p>◆類神經網路或是決策樹，皆透過增加樣本數或增加自變數之方式來提高準確度。</p>
陳文杰 ( 2004 )	高速公路交通意外與發生原因之重要關係	分類與迴歸樹 負二項迴歸	<p>◆決策樹的正確率有 58% 而負二項迴歸模式有 53%。</p> <p>◆證實資料探勘方法為可行之分析工具。</p>
王秀雯 ( 2004 )	台北市發生交通事故的受傷者為觀測者進行分析	分類與迴歸樹 多元羅吉特	<p>◆預測能力以決策樹優於多元羅吉特模式。</p> <p>◆事故嚴重程度之變數，多元羅吉特與決策樹大致相同。</p> <p>◆顯示資料探勘可以適用於運輸安全領域。</p>
尤春惠 ( 2004 )	醫療研究藥物控制	決策樹 倒傳遞網路	◆利 Bagging 及 AdaBoost 兩項分類效能提升技術可提升以 C4.5 或倒傳遞類神經網路。
郭一聰 ( 2005 )	應收帳款問題	類神經網路 決策樹	◆以決策樹與類神經網路建立的逾期預警模式，在經銷商客戶的應收帳款管理上，有不錯的預測結果。

過去應用於分析肇事之方法多採以線性迴歸模式，近年來鑑於資料之複雜性及交通事故之稀少及非負數特性與常態分配假設相違，多以卜瓦松迴歸及負二項迴歸較常採用；類神經網路具備與卜瓦松迴歸或負二項迴歸一樣，為非線性迴歸模式。迴歸模式對於變數之限制及變數間之交互作用無法表達，而類神經網路卻可以，且於各領域皆具良好之績效。因不管進行迴歸模式是或類神經網路，皆無法清楚瞭解變數之影響性排序，且無法清楚變數間之交互作用為何，而分類與迴歸樹是以以樹之型態呈現，除符合前二者之非線性迴歸模式，又具備前二者無法表達之二項優點：一為具有找出最佳變數的能力，另一為產生易於瞭解的規則。故本研究將先探討變數間之交互作用，再以分類與迴歸樹產生一組劃分好的規則判定，以期能降低易肇事地點之危險程度。



## 第三章 研究方法

本研究為了找尋交通肇事次數分析上之重要影響變數，嘗試透過類神經網路及資料探勘方法於交通肇事次數分析上，並比較與統計分析上之差異，來判斷應用之可行性。考量到資料特性為計數值資料下，故採用類神經網路中的倒傳遞網路及資料探勘中的分類與迴歸樹，統計方法上則以負二項迴歸模式來分析，進而預測並找出相關因素，以分析結果做相互比較。本章即在針對卜瓦松與負二項迴歸、神經網路及資料探勘之理論作一基礎介紹。

### 3.1 統計方法

以往迴歸模式方法分析相關肇事研究中，多以建構肇事影響因素與肇事間因果模式進行分析研究。其常見為線性迴歸模式、卜瓦松迴歸模式及負二項迴歸模式。目前國內對於肇事因果分析之研究，大多採用卜瓦松迴歸模式。近年來，亦有多數學者採用負二項迴歸模式為研究方法進行分析，且其可行性及預測能力經許多研究證實，皆較線性迴歸模式佳。因此本節將分別說明卜瓦松迴歸模式及負二項迴歸模式之建構，以及參數估計與模式解釋能力。

#### 3.1.1 卜瓦松迴歸模式(Poisson Regression Model)

##### 1. 卜瓦松迴歸模式

卜瓦松迴歸模式為一種非線性迴歸模式。適用於一段時間或空間內，事件發生為計數值的資料(Count Data)。例如：高速公路的肇事件數。

Y 必須是特定的時間或區域中計數的結果，而且 Y 屬於 Poisson 分配，型態如下：

假設 Y 是屬於特定的時間或區域中計數的結果

$$P(Y_i | \lambda_i) = (\lambda_i)^{Y_i} \exp(-\lambda_i) / Y_i! \quad (3.1)$$

$$Y=0,1,2,\dots \quad ; \quad i=1,2,\dots,n$$

$$E(Y_i) = \lambda_i \quad ; \quad V(Y_i) = \lambda_i$$

其中，個案  $i$  之  $Y$  的平均反應  $E(Y_i) = \lambda_i$ ，和個案  $i$  之預測變數  $(X_1, X_2, \dots, X_{p-1})$  及迴歸係數值  $(\beta_0, \beta_1, \dots, \beta_p)$  之間的函數關係，通常以下式表示，設成指數形式的目的是為了確保  $\lambda_i$  為正：

$$E(Y_i) = \lambda_i = \exp(x_i' \beta) \quad (3.2)$$

因此， $Y_i$  的條件平均數和變異數為：

$$E(Y_i | X_i) = \lambda_i = \exp(x_i' \beta) = V(Y_i | X_i) \quad (3.3)$$

## 2. 卜瓦松迴歸模式之參數校估

卜瓦松迴歸之參數估計方法使用最大概似估計法(Maximum Likelihood Estimation, MLE)。

$$\begin{aligned} L(\beta) &= \prod_1^n f_i(Y_i) = \prod_1^n \frac{[\exp(x_i' \beta)]^{Y_i} \exp[-\exp(x_i' \beta)]}{Y_i!} \\ &= \frac{[\prod_1^n \exp(x_i' \beta)]^{Y_i} \exp[-\sum_1^n \exp(x_i' \beta)]}{\prod_1^n Y_i!} \end{aligned}$$

對  $L(\beta)$  取對數，令  $\ln L(\beta) = l(\beta)$

$$l(\beta) = \ln L(\beta) = \sum_{i=1}^n Y_i \ln[\exp(x_i' \beta)] - \sum_{i=1}^n \exp(x_i' \beta) - \sum_{i=1}^n \ln(Y_i!)$$

讓  $l(\beta)$  極大化，尋得  $\beta_0, \beta_1, \dots, \beta_{p-1}$  的最大概似估計量。可以使用數值搜尋法或是迭代覆重加權最小平方法來尋得  $\beta_0, \beta_1, \dots, \beta_{p-1}$  的最大概似估計值。

卜瓦松迴歸被廣泛地運用到離散性的計數性資料是因為具有以下優勢：(1) 卜瓦松分配對特定時間內隨機且獨立發生的事件有合理的描述，故能夠有效地處理計數性的資料；(2) 卜瓦松分配本身雖為離散性的分配，但其參數  $\lambda$  卻為連續性的變數，故令  $\lambda$  為迴歸式不



僅考慮到外生變數的效果而且未違反因變數為連續性的基本迴歸模式假設。

雖然卜瓦松迴歸模式提供對建構計數值資料的迴歸模式，但是實際上，資料的變異數通常會大於平均數。這種情形和卜瓦松迴歸模式限制資料的平均數和變異數必須相等，互有衝突。可能使得估計出來的參數和理論上的假設會不一致。

為瞭解決卜瓦松迴歸模式的缺點，通常會讓  $\lambda_i$  符合一隨機效果的分配(Random-Effects Distribution)，模式的估計結果就能夠符合實際資料之變異數會大於平均數的情形。其中，當隨機效果的分配為 Gamma 分配時，該模式就稱為負二項迴歸(Negative Binomial Regression)。

### 3.1.2 負二項迴歸模式(Negative Binomial Regression Model)

負二項迴歸亦為非線性迴歸模式，與卜瓦松迴歸相同皆為處理特定時間內所發生的事件次數或計數值資料(Count Data)；而相異則為負二項迴歸乃透過伽瑪分配(Gamma Distribution)來描述期望的計數值資料  $\lambda$ ，使模式符合資料的平均數與變異數不相同的特性。本節將說明負二項迴歸模式之建構、模式之參數校估與模式解釋能力之判斷。

#### 1. 建構負二項迴歸模式

負二項迴歸模式可放寬卜瓦松迴歸模式的期望值與變異數之限制，在卜瓦松迴歸模式中，表示單位時間內發生次數  $\lambda_i$  的表示方式如下：

$$\lambda_i = e^{\beta X_i},$$

而負二項迴歸模式中則多了一個誤差項，表示如下：

$$\lambda_i = e^{\beta X_i + \varepsilon_i}, \text{ 可以改寫成: } \ln \lambda_i = \beta X_i + \varepsilon_i$$

其中  $\varepsilon_i$  服從伽瑪分配(Gamma Distribution)，且平均數為 1，變異數為  $\alpha$ ，因此在多了  $\varepsilon_i$  的情況下，所發生的機率為條件機率，表示如下：

$$p(n_i | \varepsilon) = \frac{\lambda_i \exp(\varepsilon_i)^{n_i} \exp(-\lambda_i \exp(\varepsilon_i))}{n_i!}, n = 0, 1, 2, 3, \dots$$

其中  $P(n_i|\varepsilon)$  表示在有誤差項之情況下，發生次數為  $n$  次之機率， $\lambda_i$  則為卜瓦松分配的期望次數，而條件機率透過積分(對  $\varepsilon_i$  積分)則可得到機率分配：

$$p(n_i) = \frac{\Gamma(\theta + n_i)}{\Gamma(\theta)n_i!} \mu_i^\theta (1 - \mu_i)^{n_i} \quad (3.4)$$

$$\text{其中 } \mu = \frac{\theta}{\theta + \lambda_i}$$

$$\theta = \frac{1}{\alpha}$$

而  $\Gamma$  表示伽瑪函數(Gamma Function)，而此機率分配之期望值與變異數分別為：

$$E\{n_i\} = \mu$$

$$\sigma^2\{n_i\} = \mu + \alpha\mu^2$$

## 2. 負二項迴歸模式之參數校估

負二項迴歸模式之參數估計的方式亦使用最大概似估計法(Maximum Likelihood Estimation, MLE)，其概似方程式如下：

$$L(\beta) = \prod_i^n P(n_i) \quad (3.5)$$

對  $L(\beta)$  取對數，使  $\log L(\beta)$  最大，找出  $\beta_0, \beta_1, \dots, \beta_{p-1}$  與  $\alpha$  的最大概似估計量。

### 3. 負二項迴歸模式的解釋能力與優缺點

負二項迴歸模式的解釋能力與卜瓦松迴歸模式相同，一樣是透過  $\rho^2$  的計算，當  $\rho^2$  越高時，會有越好的解釋能力。負二項迴歸模式其優點是若分析之資料平均數大於變異數時，則此模式可符合資料的變異數大於平均數的特性。由於此模式能符合平均數不等於變異數特性，主要是因為變異數的推倒而來，變異數表示方式為：

$$\text{Var}(n_i) = E(n_i)[1 + \alpha E(n_i)] \quad (3.6)$$

經移項後可得到：

$$\frac{\text{Var}(n_i)}{E(n_i)} = 1 + \alpha E(n_i)$$

由上式可知，當  $\alpha=0$  時，變異數  $\text{Var}(n_i)$  與期望值  $E(n_i)$  比值為 1，所以變異數等於期望值，相反的，當  $\alpha \neq 0$  時，則變異數與期望值不相等。

所以  $\alpha$  又稱做過散率(Over-Dispersion)，當變異數大於平均數時，表示資料呈現較分散的情況，另外，透過 t 檢定(T-test)，也可以判斷模式是否為負二項迴歸，透過檢定假設如下：

$$H_0 : \alpha=0$$

$$H_1 : \alpha \neq 0$$

當拒絕  $H_0$  時，則  $\alpha$  顯著不為 0，所以變異數不等於平均數，則模式為負二項迴歸模式，而當接受  $H_0$  時，則表示  $\alpha=0$ ，所以變異數與平均數相等，模式應該為卜瓦松迴歸模式。所以卜瓦松迴歸模式為負二項迴歸模式的一個特例(special case)，當  $\alpha=0$  時與卜瓦松迴歸模式相同。在應用於國道肇事次數分析上，除了解釋變數(道路幾何、使用情形、交通量等)是否顯著外， $\alpha$  值的顯著與否，關係到資料適合於卜瓦松迴歸或是負二項迴歸，若  $\alpha$  顯著不為 0，則表示此分析適合於負二項迴歸模式。

### 3.1.3 變數之彈性(Elasticities of Independent Variables)

彈性是用來評估模式中獨立變數的變化對於相依變數之影響，以本研究而言，就是獨立變數變動情形對於事故次數之影響，彈性之定義為：

$$E_{X_{ij}}^{\lambda_i} = \frac{\partial \lambda_i}{\partial X_{ij}} \cdot \frac{X_{ij}}{\lambda_i} \quad (3.7)$$

其中  $E$  表示彈性值， $X_{ij}$  為觀測值  $i$  其變數  $j$  之值， $\lambda_i$  為觀測值  $i$  期望次數，所以  $\lambda_i$  為路段  $i$  之平均事故次數， $X_{ij}$  可為路段  $i$  之交通量。由於  $\ln \lambda_i = \beta X_i + \varepsilon_i$ ，所以上式可以推導為：

$$E_{X_{ij}}^{\lambda_i} = \beta_j X_{ij}$$

其中  $\beta_j$  為  $X_{ij}$  之係數，而彈性可解釋為當獨立變數改變 1% 時，事故次數變化百分比，若交通量變數之彈性值為 1.36 時，表示當交通量增加 1% 時，該路段會增加 1.36% 的事故發生。

上式計算方式為連續變數之彈性，稱為直接彈性 (Direct Elasticity)，當變數為虛擬變數 (Indicator Variable) 時或是該變數之值只有 0 或 1 時，則要透過擬似彈性 (Pseudo-Elasticity) 來計算，其定義為：

$$E_{X_{ij}}^{\lambda_i} = \frac{\exp(\beta_j) - 1}{\exp(\beta_j)} \quad (3.8)$$

其中各變數之定義已於上所說明，相關的應用方式可參考 Shankar(1995)等學者之研究。

## 3.2 類神經網路

近年來類神經網路被廣泛應用在各個領域之中，常用來解決分類、預測、函數逼近及最佳化等問題上，由於本論文亦採用類神經網路模式作為分析之工具，故首先針對類神經網路模式之基本概念及倒傳遞網路之操作流程、其次介紹評估模式之績效的效能評估及敏感度分析之準則。

### 3.2.1 類神經網路基本概念

類神經網路是利用大量的人工神經元所組合，建構成網路架構來

模擬人類神經之功能，透過訓練資料的訓練來修正網路的參數，使整個網路模式符合訓練資料特性，並將經驗(修正的參數值)儲存在網路中。而人工神經元是生物神經元的模擬，它是將外部的輸入值，經由加總及轉換函數的作用後輸出其結果，而透過大量的人工神經元作用，可用來解決不確定性的問題或輸入及輸出間存在複雜的非線性關係問題。類神經網路的主要單元、架構與運作方法說明如下【葉怡成, 1998】:

### 1. 處理單元(Processing Elements)

類神經網路是由許多人工神經元所組成，而人工神經元即為處理單元，大多數的演算工作均在此完成，而網路中各處理單元是以適當方式連接，而形成不同網路模式。而處理單元的主要功能是將輸入資料向量經過集合、轉換等過程，產生一輸出向量成為下個處理單元的輸入值或最終的輸出值。

### 2. 類神經網路的網路架構

類神經網路是由有許多處理單元所連接而成，一般主要可分為三層，包括：

(1)輸入層：主要為放置輸入變數，神經元的個數即代表輸入變數的個數。

(2)隱藏層：主要為提供各變數間作交互作用影響之用，其中隱藏層的數目，可不只一層，可採取多層的隱藏層，在部份問題模型亦可不需隱藏層。隱藏層的主要是用來表達複雜的交互作用關係，而隱藏層中的神經元的數目亦有相似的作用。

(3)輸出層：用來輸出網路的結果，其神經元數目，依問題類型而決定。具有輸入層、隱藏層、輸出層的類神經網路又稱為多層網路，而沒有隱藏層僅有輸入層及輸出層的類神經網路則稱為單層網路，單層網路一般應用在具線性關係的問題方面，而多層網路。

### 3. 類神經網路的運作方式

因具有隱藏層則可用來處理複雜的非線性問題。

類神經網路的運作主要分為兩個過程：

(1)訓練(Training)或學習(Learning)過程：

訓練或學習之目的在決定網路的連結加權值與偏權值，此過程為類神經網路的建構階段。類神經網路可依學習方式來分類可分為監督式學習網路(Supervised Learning Network)、無監督式學習網路(Unsupervised Learning Network)、聯想式學習網路(Associate Learning Network)、最適化應用網路(Optimization Network)。

(2)模擬(Simulating)或回想(Recalling)過程：

模擬或回想之目的在決定網路的預測輸出值或驗證網路的準確度，此過程為類神經網路實際應用操作之方法。

### 3.2.2 倒傳遞網路

在眾多的類神經網路模式中，倒傳遞網路(Back-Propagation Network; BPN)模式為應用最普及的模式之一，可適用於分類與預測之問題。最早的倒傳遞類神經網路模式，是由 Werbos 於 1974 年在其博士論文中提出了隱藏層的學習演算法。至 1985 年 Parker 與 Rumelhart et al.分別發表了有關倒傳遞網路文章，使得倒傳遞網路廣為人知。倒傳遞類神經網路模式即是透過部份訓練資料，使模式將相關經驗儲存於網路各連結與處理單元之中，而不同輸入訊號，透過網路中各連結加權值(Weight)的調整後輸入至各處理單元，而處理單元再利用集成函數(Aggregation Function)與轉換函數將輸入值轉換，並產生最終之輸出值。而倒傳遞網路的運作亦具有訓練(Training)與模擬(Simulating)兩個過程，以下就敘述倒傳遞網路的操作與運用過程：

#### 1. 訓練過程

由於倒傳遞網路的訓練方法為監督式(Supervised)的網路，必須從待處理的範圍中取出一部分的變數資料及輸出參考值，作為訓練學習之範本。而有關訓練的過程是將這些訓練資料輸入倒傳遞網路

中，以最陡坡降法(Gradient Steepest Descent Method)進行連結加權值與偏權值(Bias)的反覆修正，當誤差收斂至容許值時即可完成訓練。由於連結加權值與偏權值的修正過程不但有正向傳遞的行為，同時也有倒傳遞的現象，因此稱為倒傳遞類神經網路。下式分為別連結加權值的調整方式與偏權值的調整計算方式：

$$w_{i,j}^m(k+1) = w_{i,j}^m(k) - \alpha \frac{\partial \hat{F}}{\partial w_{i,j}^m} \quad (3.9)$$

$$b_i^m(k+1) = b_i^m(k) - \alpha \frac{\partial \hat{F}}{\partial b_i^m} \quad (3.10)$$

其中， $w_{i,j}^m$ 為第  $m$  層的  $i$  單元至  $j$  單元的連結加權值；

$b_i^m$ 為第  $m$  層的  $i$  單元偏權值；

$k$  疊代次數；

$\alpha$  學習速率；

$F(x)$ 為即能量函數(Energy Function)；及  $\frac{\partial \hat{F}}{\partial w_{i,j}^m}$ ， $\frac{\partial \hat{F}}{\partial b_i^m}$ 為倒傳遞的靈敏度，可由輸入層的靈敏度倒傳遞推求各層的靈敏度，故稱為倒傳遞網路。

網路的訓練過程目的即為使量能函數最小化之過程，透過使能量達到最小化來降低理論輸出值與網路實際輸出值之差距，當每一個輸入向量作用於網路，就比較其網路輸出與其目標。而訓練演算法其功能就是要來調整網路參數以達到最小化能量(最小化誤差)，而能量函數表示方法如下式：

$$F(x) = E[e^2] = E[(t - a)^2] \quad (3.11)$$

其中， $x$  是網路權重值和偏權值向量；

$t$  為目標輸出向量；及

$a$  為推論輸出向量。

倒傳遞網路的訓練步驟：

(1)隨機設定連結加權重與各偏權值。

- (2)選擇輸入向量(p)，並將其輸入網路之輸入層中。
- (3)根據目前的連結加權值求得網路的實際輸出向量(a)。
- (4)計算實際輸出向量與理論輸出向量之差異(t-a)
- (5)調整網路的連結加權值以降低網路的能量函數

重覆步驟(2)至(5)，直到所有訓練資料均完成訓練，並使網路能量函數收斂為止。

## 2. 模擬過程

模擬的過程是利用倒傳遞類神經網路訓練完成所得的網路各連結加權值與偏權值，來進行進模擬，將輸入向量經過網路連結加權值作用，透過每個處理單元集成函數加總、偏權值修正後，再經過轉換函數的轉換後，即形成該處理單元的輸出值，再將輸出值送往下一層之連結處理單元處理，最終至輸出處理單元輸出，即完成模擬之過程。以第 n 層的第 j 個單元  $a_j^n$  的輸出值如下式：

$$a_j^n = f\left(\sum_{i,j} w_{ij} p_i + b_j\right) \quad (3.12)$$

其中， $\sum_{i,j} w_{ij} p_i$  為集成函數；

$f$  為轉換函數； $w_{ij}$  為介於第 n-1 層的第 i 個處理單元，與第 n 層的第 j 個處理單元間的連結加權值；及  $b_j$  為第 n 層的第 j 個處理單元偏權值。常用之轉換函數( $f$ )大多為非線性之函數，包括有雙彎曲函數(Sigmoid Function)及雙彎曲線正切函數(Hyperbolic Tangent Function)。其函數特徵如下：

雙彎曲線函數(Sigmoid Function)

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3.13)$$

其值介於 0 至 1 之間，當 x 值趨於無窮大時，函數值趨近於 1；當 x 趨於負無窮大時，函數值趨近於 0。

雙彎曲線正切函數(Hyperbolic Tangent Function)



$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3.14)$$

其值介於-1 至 1 之間，當  $x$  值趨近正無窮大時，函數值趨近於 1；當  $x$  值趨近負無窮大時，函數趨近於-1。

### 3.2.3 效能評估

為瞭解類神經網路訓練的成果，在訓練網路前必需將樣本隨機分類成訓練組資料與測試組資料，做為驗證類神經網路效能之用。以訓練組資料來訓練類神經路，當訓練完畢時，再以測試組資料來驗證網路性能是否良好。網路的誤差表示方法有下列兩種：

#### 1. 均方根誤差(Mean Square Negligence, MSE)

函數型問題之網路誤差可用均方根誤差來表示，下式為均方根誤差之定義：

$$F = \frac{1}{N} \sum_{i=1}^N (t_i - a_i)^2 \quad (3.15)$$

其中， $F$  為均方根誤差；

$N$  為範例數目；

$t_i$  為第  $i$  個目標輸出值；及

$a_i$  為第  $i$  個推論輸出值。

#### 2. 正確率

分類型問題之網路誤差可用正確率或誤判率來表示，下式正確率之定義：

$$\text{正確率} = \frac{\text{正確判斷的測試範例數}}{\text{測試範例總數}}$$

均方根誤差與正確率的變化情形大致上呈現固定，一般而言均方根誤差愈小相對的其正確率也就愈高。

### 3.2.4 敏感度分析

此分析之主要目的是藉由敏感度擷取來確定每個變數在網路模式中的影響程度，因此可使用敏感度擷取估計每個輸入變數的影響值，再剔除影響較小的輸入變數，將可提高網路之效能。敏感度擷取模式可從加權矩陣式知識庫中，擷取網路特定輸入層處理單元與特定輸出層處理單元間之敏感因子。敏感度正值越大表示兩者間之正相關越大，敏感度負值越大表示兩者間之負相關越大。敏感度  $S_{ij}$  定義如下【Baba 等，1991】：

設  $X_i$  為輸入層第  $i$  個處理單元輸入值

$H_k$  為隱藏層第  $k$  個處理單元之輸出值

$Y_j$  為輸出層第  $j$  個處理單元之推論輸出值

$net_j$  為輸出層第  $j$  個處理單元之集成函數

$net_k$  為隱藏層第  $k$  個處理單元之集成函數。

則  $X_i$  對  $Y_j$  透過隱藏層第  $k$  個處理單元之敏感度  $(S_{ij})^k$  如下所示：

$$\begin{aligned} (S_{ij})^k &= \left[ \frac{\partial Y_j}{\partial X_i} \right]^k = \left[ \frac{\partial Y_j}{\partial net_j} \right] \cdot \left[ \frac{\partial net_j}{\partial H_k} \right] \cdot \left[ \frac{\partial H_k}{\partial net_k} \right] \cdot \left[ \frac{\partial net_k}{\partial X_i} \right] \\ &= f'(net_j) \cdot W_{ij} \cdot f'(net_k) \cdot W_{ik} \end{aligned} \quad (3.16)$$

若假設  $f'(net_j)$  與  $f'(net_k)$  為常數，忽略不計，則可得敏感度  $S_{ij}$ ，如下述公式：

$$S_{ij} = \sum_k (S_{ij})^k = \sum_k W_{kj} \cdot W_{ik} \quad (3.17)$$

其中  $W_{kj}$  為第  $k$  個隱藏層單元與第  $j$  個輸出層單元間之連結權重

$W_{ik}$  為第  $i$  個輸入層單元與第  $k$  個隱藏層單元間之連結權重

### 3.3 資料探勘

本節主要針對資料探勘概念、分類和迴歸樹、Gini 與 Twoing 法則及變數重要性分別介紹，藉以瞭解資料探勘含義，並且從許多的資訊技術中找出分析本論文之研究方法，以探討影響交通事故嚴重程度之因素及影響程度。

### 3.3.1 資料探勘概念

#### 1. 資料探勘定義

資料探勘(Data Mining)的定義，簡單而言，是指從資料中萃取有效的、新穎的、具潛在效益的資訊之細瑣過程，其最終目標是瞭解資料的型樣(Pattern)。其主要含義如下列所述【Fayyad, 1996】：

- (1).資料在資料庫中代表一組事據(Facts)。
- (2).型樣是一個存在資料組中表達描述事據的語言。
- (3).資料探勘程序有許多步驟，包含資料前處理、型樣搜尋、知識評估及修改精煉。
- (4).找尋到的型樣有一定程度的根據。
- (5).至少對於系統來說型樣是新穎的。
- (6).型樣對於決策行動有潛在的助益。
- (7).資料探勘的目標是使型樣可讓人易於理解。

實際上資料探勘只是KDD(Knowledge Discovery in Databases)找尋隱藏在資料中有用的知識和型樣程序中的一個步驟，而整個KDD的流程則包括選擇資料、資料轉換、資料前處理、資料探勘、評估與解釋等程序，但現在只要提到資料探勘，就是指代表同義的KDD。KDD的目的是從資料中瞭解其特性並發現由此產生的規則，其規則的屬性必須具有關聯(Relevance)、新穎(Novelty)、有效(Efficacy)、簡易(Simplicity)的特性，也就是說所找到的規則必須與解決的問題有關聯性，並具有非一般、非普通的特性，且能簡單有效地應用在解決問題上。資料探勘的資料來源不限於資料庫，也包括其它類型的資訊儲存形式，及範圍廣大的資料倉儲(Data Warehouses)。資料探勘主要是運用統計模式、數學演算法、機器學習方法等工具，其不只包括蒐集資料和管理資料，也包含分析和預測的功能。

#### 2. 資料探勘的功能

Berry and Linoff 將資料探勘歸納為五種功能，包括【 Berry, 1997 】:

#### (1)分類(Classification)

分類是資料探勘作業中最普遍的一種，是按照分析對象的屬性分門別類加以定義，再將其指派至一現有類別中，建立類組(Class)。例如：將信用狀況區分為高度風險、中度風險及低度風險申請者、保險詐騙事件的判斷、新聞報導的分類、顧客反應預測，可能有類似的消費行為等。使用的技術有決策樹(Decision Tree)、記憶基礎理解(Memory-Based Reasoning, MBR)、連結分析(Link Analysis)等方法。

#### (2)推估(Estimation)

推估是善於處理連續性的數值，根據既有連續數值之相關屬性資料，以獲致某一屬性未知之值。例如：按照信用申請者之教育程度、行為別來推估其信用卡消費量。其使用的技術有統計方法上之相關分析(Correlation Analysis)、迴歸分析(Regression Analysis)及類神經網路(Artificial Neural Networks)等方法。

#### (3)預測(Prediction)

預測是去推估未來的數值以及趨勢。根據屬性之過去觀察值來推估其屬性未來值。例如：由顧客過去之刷卡消費量預測其未來之刷卡消費量，那些顧客會在特定時間內流失、預測那些商品會被同時購買之行為等。其使用之技術計有迴歸分析(Regression Analysis)、時間數列分析(Time Series)、決策樹(Decision Tree)、記憶基礎理解(Memory-Based Reasoning, MBR)及類神經網路(Artificial Neural Networks)等方法。

#### (4)關聯分組(Affinity Grouping)

關聯分組的功能是去發覺那些事物總是同時發生，以決定那些相關事物應該放在一起。例如：超市中相關之盥洗用具，放在同一貨架上，在行銷系統上也可以用來確立交叉銷售(Cross Selling)的基

礎，以藉此設計促銷行動方案。其使用之技術為購物籃分析法(Market Basket Analysis, MBA)。

#### (5) 集群(Clustering)

集群將一群異質的群體區隔為同質性較高的群組(Clusters)。集群相當於行銷術語中的區隔化(Segmentation)，但在集群作業中，並不需要事先定義好該如何分類，資料是靠自己的相近性來群聚在一起的。其使用之技術計有 K-Means 法。

相較於以上的分類，Ham and Kamber 則將資料探勘的功能分為六種，除了原本的分類(Classification)、集群(Clustering)相同外，還加上下列四項：

##### ①. 迴歸(Regression)

根據既有連續性數值之相關屬性資料，以獲得某一屬性未知之值。

##### ②. 規則衍生 ( Rule Generation )

從資料裏挖掘或產生規則，如關聯法則的挖掘意指在不同的因果要素中找出其關聯屬性。

##### ③. 總合或濃縮 ( Summarization or Condensation )

提供資料組壓縮整理的功能，特別是針對多媒體資料。

##### ④. 序列分析(Sequence Analysis)

可用來建立序列型樣的模式，如時間序列分析(Time Series)、基因序列(Gene Sequences)分析等。

### 3. 資料探勘的技術

資料探勘技術除了統計學(Statistics)之外，還包括基因演算法(Genetic Algorithm)、類神經網路(Artificial Neural Networks)、記憶基礎推理法(Memory-Based Reasoning, MBR)、購物籃分析(Market Basket Analysis, MBA)、決策樹(Decision Tree)、群集偵測(Cluster Detection)、連結分析(Link Analysis)等等技術。常見的資料探勘技術

依其應用範圍列表如下【Cabeña, 1997】:

表 3.1 資料探勘常用的技術

應用範圍	描述	演算法
預測模型	提供一個特定輸入值，預測一個連續或離散輸出值	C5.0, decision trees, Bacon's algorithm, regression, moving averages, case-based reasoning, CART
集群	自動集群同質資料群	SOM, K-means
關聯分析	找出交易項目間的關聯性	A priori algorithm
違規偵測	察覺非法者	Regression

若就資料探勘探尋的過程來看，其主要活動包含三類程序，如圖 3.1：

(1)發現(Discovery)—從資料庫中找尋未知而隱藏的型樣之過程。

(2)預測建模(Predictive Modeling)—從資料庫中發現型樣並使用它預測未來之過程。

(3)辯證分析(Forensic Analysis)—應用萃取出型樣去尋找異常或奇特的資料元素之過程。

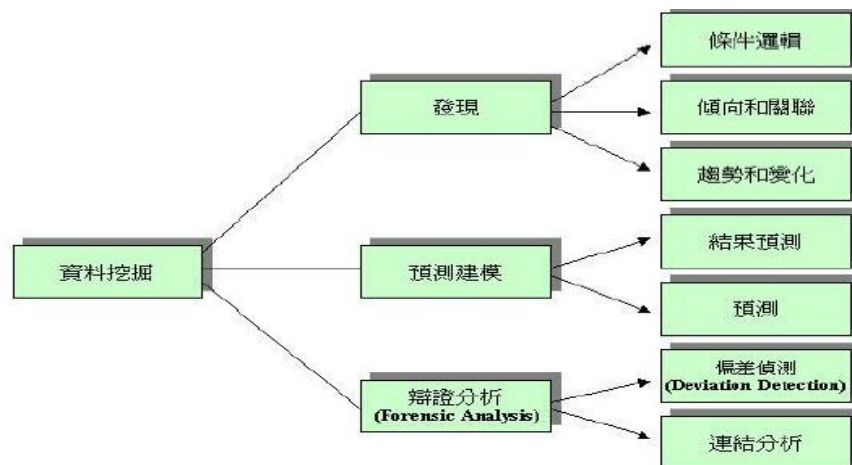


圖 3.1 資料探勘程序的分類

若就資料探勘技術的分類來看，可分為如圖 3.2：

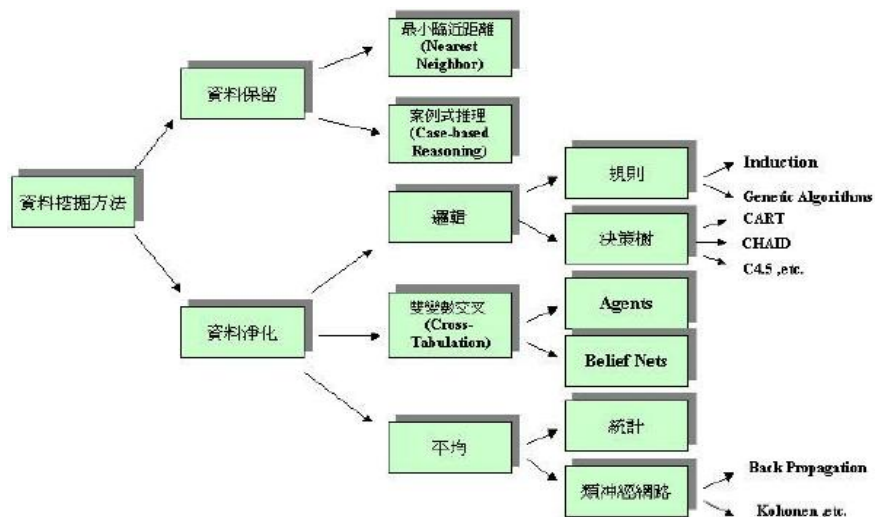


圖 3.2 資料探勘技術的分類

資料來源：【 <http://www.dmreview.com/> 】

過去文獻針對決策樹方法分析有 C4.5、CHAID、ID3、分類與迴歸樹等方法。Breiman, et al., 1984 以分類與迴歸樹 (Classification and Regression Trees, CART) 進行分析，發現決策樹是可以歸納出事例的一些規則，以單一變量分割而二元樹，其優點在於：

1. 它是一個非參數的分類守則，而且不需要事先的假設規則，所以由此樹可以很容易讓使用者得到所要的資訊。
2. 分類與迴歸樹是不需要停止規則，而只需要將樹發展完全再由修剪方式獲得最適樹。
3. 分類與迴歸樹是以節點之純度進行分割，其最後分割之終端節點只存在一個類別。

由上述三點所言，本論文則針對決策樹作為本論文方法。

### 3.3.2 分類與迴歸樹

分類與迴歸樹(Classification and Regression Trees, CART)在 1984 年由 Breiman, Friedman, Olshen 和 Stone(Breiman, et al., 1984)四人所提出，應用於解決分類的問題上，是一種簡單且有效的無母數工具。其基本原理是使用二元遞迴分割(Binary Recursive Partitoning)過程來分析龐大的資料集，CART 會依照預測變數與其相對的各項指標，將既有的訓練樣本劃分成數個已知的類別，並將其劃分程序彙總成一連串的規則(Rule)。

然而，早在 1960 年代開始便有許多文獻以樹狀結構方式進行資料分析，只不過早期的分析工具僅能處理連續型的變數，對於間斷型變數的分析，因為無法同步處理，所以難達到理想的成效。爾後，CART 的發展，解決前述的問題，成為可以同步處理連續型及間斷型資料的分析工具。再者傳統決策樹的推導過程，會事先設定決策樹的層數或節點數，一旦滿足限制式，演算的過程便會停止；但在 CART 的運算過程中，會產生包含許多節點的最大樹狀結構(此時該樹的末端節點數目為最大值)，然後再依照所挑選的準則進行決策樹的修剪(Pruning Procedure)，直到滿足停止修剪的準則為止。CART 的運算雖然會花費較長的時間，但能避免資訊隱藏的缺點；此外 CART 將分析結果以決策樹呈現，得以直接看出重要變數及其區隔效果，能使研究結果更加清楚易懂。



整體而言，分類與迴歸樹的分析流程大致可分為三個步驟：首先，將樣本資料劃分為訓練樣本與測試樣本，並根據劃分規則使用訓練樣本以建構最大數目的決策樹；其次，依照修剪準則，從樹的底部向上修剪，並產生子樹群(Subtrees)，直到修剪準則被滿足為止；最後，使用測試樣本進行交叉驗證(Cross Validation)並在子樹群中挑選一最佳樹狀結構。待樹狀結構建置完成後，CART 將會產生一連串的分類規則，該規則可幫助決策者在輸入新樣本資料後，正確的將新樣本劃分到適當的分類群組中。其進行步驟詳述如下：

### 1. 建構最大樹狀結構

CART 的基本原理是藉由二元遞迴分割過程以處理資料，因此面對每一個節點時，將針對資料向下劃分為兩個資料子集，直到停止條件被滿足為止。而 CART 模式建構過程之所以能夠較有效率，則是因為其進行各節點切割工作時，並不考慮該切割點對整體表現的影響。而切割點的選取準則，則是每個節點的異質性(Heterogeneity)高低，而用於評估節點異質性評估的指標有三種，分別是 Twoing，Entropy 及 Gini，其中 Gini Index 較為普遍。Gini Index 的計算方式如下：

$$Gini(t) = 1 - \sum_i p_i^2 \quad (3.18)$$

其中  $t$  為已知節點  $t$ ； $P_i$  為第  $t$  個節點中第  $i$  個類別的機率。

當 Gini 值越小時代表在該切割點所產生的子節點中組成份子越單純，即該切割點判別不同類別差異的能力越好，因此所造成之誤置率也越低。當 CART 要決定某一末端節點所屬類別時，通常以計算該末端節點中各類別所包含的資料數目多寡為準則，選取個數最多的類別作為該末端節點的類別。又因為每個節點都可能是末端節點，因此，進一步計算每個節點的誤判成本，作為決定該節點是否需要繼續細分下去的準則。當樹的建構程序完成後，則此樹狀結構稱之為最大樹(Maximal Tree)。

## 2. 修剪樹狀結構並產生子樹群

找出最大樹結構後，CART 便開始修剪樹狀結構以降低誤判率，並找出最適當的決策樹，決定最適決策樹所依據的準則為誤判率 (Negligence Rate) 或誤判成本 (Negligence Cost)。CART 會計算最大樹與所有子樹的誤判率，而最適的決策樹是具有最低誤判成本或最低誤判率的樹狀結構。而在決策樹修剪過程所考量的則是重代估計值 (Resubstitution Estimate)。當重代估計值越小時，代表末端節點個數越多，即樹狀結構中的切割點越多，樹狀結構也越大。CART 會依序計算每個節點的重代估計值，由下往上逐步修剪樹狀結構，直到根節點時才停止修剪的步驟。在計算重代估計值時，亦可能重複計算到已被刪除的節點，而降低了模式建構的效率，因此，計算重代估計值時也將同時考慮誤判成本，以增進修剪的效率。

## 3. 挑選最佳樹狀結構

當所有的樹狀結構都一一呈現後，運用測試樣本來計算誤判率，即將測試樣本代入所有的樹狀結構並計算每個結構的誤判率，最後再選取最小誤判率的樹狀結構為最佳樹。

### 3.3.3 Gini 分類法

Gini 分類法在 CART 之中為主要的不純度指標 (Impurity Index)，其定義為：

$$i(t) = \sum_{j=1} p(j|t) \quad (3.19)$$

表示在節點  $t$  中，每 2 種不相同類別的樣本比率相成之總和，其可以解釋為當節點  $t$  中對類別  $i$  而言，其佔的比率  $p(j|t)$  就是  $t$  節點中類別  $i$  的純度，但是若節點  $t$  中含有其他類別，像是類別  $j$ ，則節點  $t$  就不會是最純的節點，所以其純度會被降低，也就是 2 個類別的相乘，這些所有被降低純度的總和就是節點  $t$  的不純度指標。

### 3.3.4 Twoing 分類法

Twoing 分類的方式，其原先的概念並非要計算出該節點的不純度，而是計算父節點與子節點間最大的不純度改善值，目的在將所有的類別區分成兩大類，所以稱做 Twoing(Breiman, 1998)。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(superclasses)如

$$C_1=\{j_1, \dots, j_n\}$$

$$C_2=C_1-C$$

Twoing 進行分割時，將類別  $C$  分成  $C_1$  與  $C_2$  的動作視為一個 2 類別問題(Two-Class Problem)，所以不純度改善值  $\Delta i(s,t)$  的大小，受到所設定的類別  $C_1$  影響，因此將不純度改善值改寫成  $\Delta i(s,t,C_1)$ 。

此時分割某類別時，其分割條件並不會為原來的  $s$ ，因此分割條件改寫為  $s^*(C_1)$ ，表示在類別  $C_1$  的情況下所做的分割。

Twoing 定義的不純度函數為：

$$\Phi(s,t)=\frac{P_L P_R}{4} \left[ \sum_j |p(j|t_L) - p(j|t_R)| \right]^2 \quad (3.20)$$

Twoing 在分支 (Splitting) 時，並不是評估單一節點的不純度，而是父節點與子節點間，分類上能改進的最大不純度值，而這樣的分類方式，與 ID3 演算法的 Entropy 值相當類似。

### 3.3.5 重要性變數(Importance Variable)

將具有貢獻的變數建構出一顆樹，其節點上的變數則決策樹會以排序方式將主要變數、次要變數篩選出來，從這些變數之中可以瞭解哪些變數著實會影響其研究主題，然而變數的重要性是可以在樹中得出資料的隱藏含義，因此重要性變的排序法則是透過最主要的分割者(Main Splitter)或是代理者(Surrogate)，在一棵樹中將某特定變數(Particular Predictor)交叉加總所有節點的改善分數(Improvement Scores)，其改善分數指的是在樹建構的過程中各節點從所有變數

(Competitor)選擇出主要的分割者或是代理者之節點權重(改善子節點的不純度)。因此節點之貢獻度會關於變數的重要性，則改善分數是依據節點求算出改善值，其改善值表示為：

$$\text{importance\_contribution\_node\_i} = \text{improvement} \quad (3.21)$$

當主要變數由第  $n$  個代理者代替，其公式表示為：

$$\text{importance\_contribution\_node\_i} = p^n * \text{improvement} \quad (3.22)$$

其中， $p$  是代理者改善權重(Surrogate Improvement Weight)，介於 0~1 間，預設值通當是 1。此公式定義出代理者之貢獻度是較少相關於主要分割者之變數的改善值。

### 3.4 各模式之優缺點比較

本節將對上述之研究方法統整出各預測模式優缺點比較，如下：

#### 1. 類神經網路之優缺點

類神經網路的優點有：

- (1).可以建構非線性的模型，並可以處理的問題很廣。
- (2).縱使在很複雜的領域，還是可以得到很好的結果。
- (3).能夠處理邏輯、數值、有序分類、無序分類變數作輸入，適應性強。
- (4).具有良好的推廣性，對於未知的輸入亦可得到正確的輸出。
- (5).應用在很多套裝軟體中。

類神經網路的缺點有：

- (1).類神經網路因其中間變數(隱藏層)具有很大的彈性，可由使用者自行設定層數，因此網路優化的工作相當費時。類神經網路因具大量可調參數(連結加權值與門檻值)，因此容易發生學習過度的現象，即網路對訓練範例的誤差很小，對測試範例的誤差卻很大的現象。

象。

(2).所得到的解有無限多組，無法得知哪一組的解為最佳解。

(3).訓練的過程中無法得知需要多少神經元個數，太多或太少的神經元均會影響系統的準確性，因此往往需以試誤的方式得到適當的神經元個數。

(4).無法套用公式之方式來應用

## 2. 分類與迴歸樹之優缺點

分類與迴歸樹的優點有：

(1).具有找出最佳變數的能力：不但可以知道各個變數對於結果之影響程度，更可以提供其他模式對於變數挑選的參考，本研究以分類與迴歸樹演算法之技術，正是基於此項考慮。

(2).可以產生易於瞭解的規則：即為使用本技術最大的理由，我們希望藉由法則的產生，提供決策者規劃路口能有更直覺的瞭解，而不是一堆數字，方便他迅速做出決策。

(3).建模與預測迅速：在本研究三個建模技術中，分類與迴歸樹不論在建構或預測上，均比其他二類建模技術迅速甚多。

(4).可以處理數值型及類別型變數：對於本研究某些類別型的變數，如工程因素變數，不需轉換即可使用。

分類與迴歸樹的缺點有：

(1).對於預估型的工作，如連續變數的預測，較不適宜。

(2).當類別太多時容易犯錯。

(3).多個輸出屬性是不被允許，輸出屬性必須被分類。

(4).較不穩定，當屬性的選擇影響所有子樹時，在訓練資料中的些微變化，將會導致在樹中各個選擇點有不同的屬性選擇。

### 3. 類神經網路與迴歸分析之比較

類神經網路與迴歸分析相比其相同之處有：

- (1).都是依數據建模型。
- (2).都有可調參數。在迴歸分析中為迴歸係數  $B_0, B_1, B_2, \dots$  等；在類神經網路中為網路的連結加權值與門檻值。

類神經網路與迴歸分析相比其相異之處有：

- (1).迴歸分析為線性模式；類神經網路為非線性模式。
- (2).迴歸分析無法表達輸入變數間之交互作用；類神經網路可以表達輸入變數間之交互作用。
- (3).迴歸分析的可調參數(迴歸係數)的數目是固定的；類神經網路因為其中間變數(即隱藏層)可以是一層或二層，數目也可設定為任意數目，因此其可調參數(連結加權值與門檻值)的數目是可變的，而且經常遠多於迴歸分析。
- (4).迴歸分析用解矩陣的方式一次解出迴歸係數；類神經網路因為是非線性模式，要用替代方式逼近最佳的連結加權值與門檻值。
- (5).迴歸分析的迴歸係數有唯一解；類神經網路因為是非線性模式，其連結加權值與門檻值無唯一解，也很難證明所得的解是最佳的一組解。
- (6).迴歸分析的變數限為連續值之變數；類神經網路不受此限制，數值、分類變數均可。

### 4. 類神經網路與分類與迴歸樹之比較

類神經網路與分類與迴歸樹相比其相同之處有：

- (1).兼具處理非線性問題，且輸入變數皆可收受連續及分類變數。
- (2).皆有分類及預測功能。

(3).可透過敏感度分析瞭解變數於模式之影響程度；分類與迴歸樹以變數重要性表示變數之影響程度。

(4).應用領域皆廣泛。目前有醫學、農業、企業、金融及交通領域等。

類神經網路與分類與迴歸樹相比其相異之處有：

(1).類神經網路以仿效生物神經概念發展，利用大量平行而分散的相連人工神經元，透過層次將結果輸出；分類與迴歸樹以樹的概念，透過規則性，將資料作詳細的分類。

(2).類神經網路做歸納或預測時是屬於「黑箱式」(Black Box)作業；分類與迴歸樹則根據「If-Then」方式，將不同屬性作分門別類。

(3).類神經網路運算速度慢；分類與迴歸樹運算速度較快。

### 3.5 小結

由於肇事次數屬於計數性資料，過去常使用之統計模式為卜瓦松迴歸或負二項迴歸，若資料的期望值與變異數不相等時，可透過負二項迴歸中  $\alpha$  參數是否顯著來判斷資料適合的模式。在模式中，解釋變數部分可以透過係數的正負符號來解釋對於肇事次數發生機率的增減，處理變數對模式之影響程度之方法有許多，諸如：文獻回顧法、因素分析、彈性分析、多變量分析、主成份分析及敏感度分析。本研究從文獻回顧法獲得，針對交通事故之影響因素多採以因素分析、彈性分析及敏感度分析。茲因本研究負二項迴歸所解出之迴歸係數為  $B_0, B_1, B_2, \dots$  等，為一定數目，且所得之結果為機率期望值；類神經網路於隱藏層層數及神經元數目皆可任意設定，使連結加權值與門檻值亦任意變動，故由加權矩陣式知識庫中，擷取網路特定輸入層處理單元與特定輸出層處理單元間之敏感因子。經文獻回顧法及本研究限制下，選擇較適合之方法。是故本研究為了瞭解路口交通事故之影響因素，針對類神經網路與負二項迴歸模式分別進行敏感度及彈性分析。

CART 包含迴歸樹與決策樹，由於迴歸樹較適合連續型數字變數

的分類，而路口肇事次數屬於計數型資料，因此不適合應用迴歸樹，故採決策樹進行分析。決策樹中又包含 Gini 與 Twoing 兩種分類法，由於無法明確指出肇事次數分析較適合於何種分類法，但軟體手冊中建議當分類的類別數在 4 到 9 個類別時，Twoing 有可能產生較佳的分類法則，因此仍需實證分析時，經由不同的分類法，比較不同的結果來判斷最適合的決策樹。並可透過變數重要度來了解該變數對於決策樹是否有貢獻，利用此重要度的大小來判斷變數與肇事次數之間的關係。





## 第四章 變數選取與資料蒐集分析

本研究主要探討道路環境與交通工程設施對路口交通事故發生之重要潛在因素之關係，所需之研究資料除了應具曾進行過交通流量調查之路口之交通事故資料外，尚包含道路幾何設計、交通量資料等兩種資料。故本章將闡述採用行向組合與模式相關變數之原因，其次說明資料之蒐集。

### 4.1 行向組合【18】

對交通肇事而言，所牽涉的交通問題，可分為「人」、「車」、「路」三大要素，而其發生必受到各種因素之交互影響，而構成交通肇事之發生。因三者有密切關聯，故在研究與分析交通問題時，對此三者的條件和限制，皆需徹底瞭解。

觀察肇事資料或是實際進行路口觀測，可以發現某一路口、路段或區域，在車輛、用路人及無其他特殊條件下，容易發生特定型態的事故，故可推測此一地點之道路特徵與肇事間具有某種關聯，亦可假設該路口、路段或區域，具有某種型式、特徵或組合，致使該路口產生交通危安因子，導致交通事故發生。

對於肇事地點之工程因素，有些與整個路口有關，如路口面積之大小或路口之夜照明程度；有些則與各臨近路段相關，如一臨近路段之道路寬度及是否設有左轉專用車道或左轉專用號誌等。本研究即在探討工程因素與肇事發生次數間之的關係與影響程度。本研究分析路口交通事故之影響因素，而將路口進行分解，產生不同行向之組合，建立交通事故次數預測模式。

各臨近路段進入路口時，其所面對之道路環境均有所不同，本研究假設這些相異的道路環境對肇事有所影響。因此將整個路口之肇事，依照其進入路口之上游路段區分出來，定義所謂『行向』為：「車輛進入路口之上游路段」。如此一個三岔路口便有三個行向，四岔路口則有四個行向。

事故類型中，大部分事故均有兩輛或多車輛涉及。各肇事車輛均有其行向，因此形成了不同行向組合。在行向組合之計算上，將一路口之所有行向任取兩者，由於同一行向在組合中，可被同時選取，如此一個  $N$  岔路口便有  $\frac{N(N+1)}{2}$  種組合。若以一四岔路口為例，便有  $\frac{4(4+1)}{2}$ ，共 10 種組合，如圖 4-1 所示：

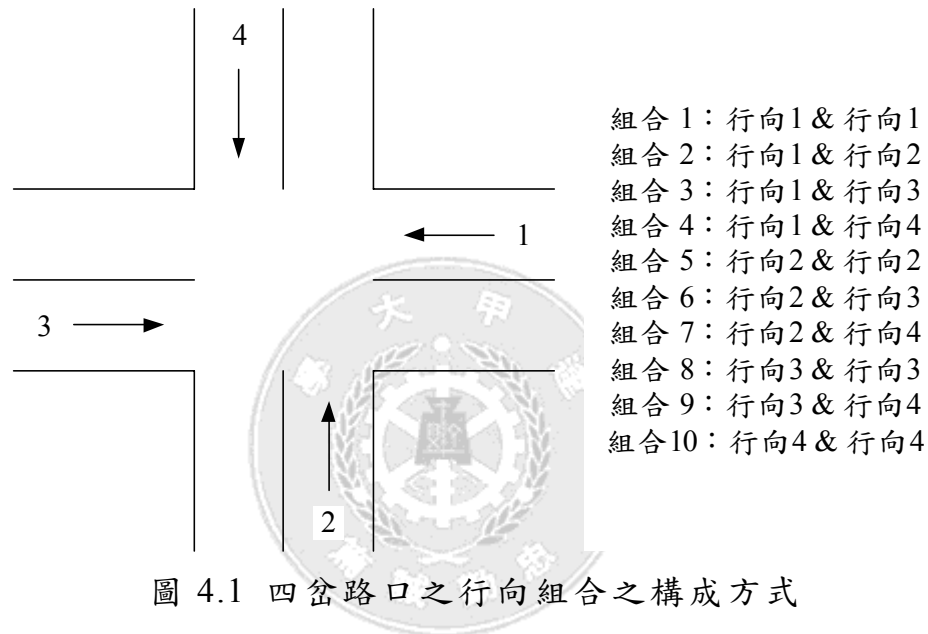


圖 4.1 四岔路口之行向組合之構成方式

本研究採用行向組合作為研究對象，原因如下：

在各種車流交會的路口中，包括標準的十字路口、三岔的 T 字路口、超過四岔的多岔路口、畸形路口以及圓環等各種不同的路口。若以路口整體之特徵加以解釋，並無法突顯各行向組合之異同，故將路口分解為行向組合，如此一來不同路口間之差異便可消除，形成一致化的分析對象。

若以個別肇事而言，每件肇事均各自有其原因，將其合併至各行向組合中，可使彼此間之差異性消失。然而在特定地點之肇事工程因素研究中，乃是將其他導致肇事之人與車輛等因素，視為各地點之均有相似的分佈。例如通過一路口之車輛駕駛人之性別與年齡分佈，便

假設與其他路口並無不同。本研究在時間、人力各方面之考量下，採用行向組合之方式對肇事工程因素進行分析。

本研究以每發生一件交通事故依照其行向記錄之，為避免產生肇事之輸入資料發生大量無肇事之行向組合，對模式產生偏誤之影響，故將道路交通事故調查報告表上各種交通事故種類合併，僅考量行向組合中，事故所發生的次數。而於處理超過兩輛車之多車肇事，則由道路交通事故調查報告表上之肇事現場圖加以研判，以造成肇事之主要車輛為其行向。例如一十字路口東西向為紅燈，西向車輛闖紅燈撞上南向車輛，而涉及東向停等車輛，則記為西向與南向之肇事。依據「道路交通事故處理實務」，當事人之順序首先依照肇事責任(過失)排列，故一般而言多車事故多半可假設前兩車之資料為造成肇事之主要車輛，並可利用觀察肇事現場圖加以驗證。而亦有可能自上述無法判斷是否該車輛為形成肇事之主要車輛，若遇有此種情況則捨棄此筆資料，例如：肇事車輛逃逸，無法判斷其行向。

## 4.2 共線性與變數選取

本研究沿襲楊憲國君所撰「以類神經網路推導路口交通事故成因與改善之研究」中，所歸納出影響交通事故發生相關之工程因素與交通動態變數，其中所選擇的變數約有 29 個，包含 9 個類別變數與 20 個連續變數，然其所引用之模組變數繁多，在實際運用時將產生難度，因此本研究透過統計之共線性篩減模組變數，擷取較適之變數組合，以增加變數代表性及模式可用性。

### 4.2.1 共線性之概念

在迴歸分析中，經常可以發現一種現象，即若兩個自變數之間相

關性高，對於一變數的相關亦高，但執行迴歸分析後，可能僅剩其中一個自變數對一變數具有預測力，另一個變數的預測力則無故消失。此現象稱為高度多元共線性(Multi-Collinearity)。

所謂多元共線性係指一組變項間具有高度之相關，由於這些變項間之相關太高，在計算時會受到彼此高度共線的影響，而造成估計不穩定，或某些數學估計程序失敗之現象(例如分母項為 0 的現象)。因此，為了瞭解變數間是否為線性關係(Linearity)或是否為共線性(Collinearity)之情況，基本作法是看雙變項之相關矩陣。倘若因變數與自變數間之關係很弱或比自變數間之相關弱，應質疑所設定之多元迴歸模式是否適當。檢視自變數與因變數間是否為線性關係的基本作法是看雙變項間之散佈圖(Scatter Plot)。而發現共線性問題存在，基本修正辦法有底下幾點方法：

1. 去除造成共線性的變數(有副作用，因為只要不是百分之百線性相依，去除其中一個可能影響變數之解釋能力)。
2. 蒐集更多資料(增加  $n$  的個數有時會減弱解釋變數向量間的線性相依關係，當資料不夠時，會錯覺一個變數是另一個的代理者)。
3. 接受共線性問題的存在，因為不影響變數之適用性。

本研究因應解決實際運用時的難度，故以去除造成共線性的變數之方法來擷取變數。

#### 4.2.2 變數選取

路口一旦發生事故，決策者或是警察應清楚判斷是否幾何設計及規劃之道路環境不當所引起。而交通事故與駕駛行為、幾何、號誌及交通現況等因素交互影響且錯綜複雜，而各因素對事故之影響程度亦不易界定。所以利用彼此間相關程度先進行顯著變數篩檢，如表 4.1 所示。

表 4.1 變數相關分析表

	快車道 個數 1	快車道 個數 2	道路寬 度 1	道路寬 度 2	左轉用 號誌 1	左轉用 號誌 2
快車道寬度 1	0.9364					
快車道寬度 2		0.9270				
快車道以外寬度 1			0.8898	0.7846		
快車道以外寬度 2			0.7576	0.8982		
時相複雜程度 1					0.7911	
時相複雜程度 2						0.7916
左轉秒數 1					0.8400	
左轉秒數 2						0.8465

從上述分析來看，本研究選取變數與變數間之共線性超過 0.75 以上，如快車道寬度與快車道個數、快車道以外寬度與道路寬度、時相複雜程度與左轉用號誌及左轉秒數與左轉用號誌。為了不受共線性問題之影響，造成預測力降低，將其逐步回歸分析，比較整體  $R^2$  提升或降低作依準。若變數讓  $R^2$  值降低，則將其剔除；反之，則保留。如原  $R^2$  值為 0.473，加入快車道寬度變數後， $R^2$  值為 0.442，因  $R^2$  值降低，故剔除此變數；放入快車道個數，則  $R^2$  值變為 0.513，故將變數快車道個數保留；而時相複雜程度與左轉用號誌，茲因兩變數皆能讓整體之  $R^2$  值提升，分別為 0.497 與 0.484，又因所提升之差距不大，故將此兩變數保留。經整理後，選取選取 20 個變數，分別為快車道個數、左轉專用道、道路寬度、中央分隔設施型式、快慢車道分隔型式、號誌個數、左轉專用號誌、時相複雜程度、時制週期、綠燈秒數、速限或警告標誌、是否位於坡道、十公尺內有障礙、車道被佔用程度、路口總交通量、行向交通量、尖峰小時係數、左轉車輛比、右轉車輛比、機車比。

而本研究經實際現勘及蒐集，發現路口是否位於坡道與尖峰小時係數兩個變數在拾得上不易取得，故本研究將其去除，最終採取 18 個變數。

### 4.3 相關變數之說明

本研究之變數說明彙整如下所述：

#### 4.3.1 工程因素變數

##### 1.快車道個數

考量車道數對於能負荷之車流量影響，以其各行向之快車道數為輸入單位。

##### 2.左轉專用道

隨各路口車流方向組成之差異。若一路口有大量的轉彎車輛，於綠燈時，左轉車輛與對向直行車輛易發生衝突，因此可能產生待轉車輛佔用直行車道，而對直行車流產生影響。因此考慮有無專供車輛待轉之左轉車道，而於本研究中分為：

代號	情況
0	無左轉車道
1	有 1 條左轉車道
2	有 2 條左轉車道

##### 3.道路寬度

道路本身之寬度將影響其所能負荷之車流量。車流擁擠時，除前後間距縮小之外，亦容易有車輛不依循車道劃分或擠入慢車道行駛，導致事故發生，而此時道路寬度較個別車道寬度更能反映流量與容量之間的關係。

##### 4.中央分隔型式

道路非單行道者，於雙向行車間有所分隔，依據其跨越之因難度，

有數種不同的設置。許多文獻中指出道路槽化(含中央分隔以及快慢分隔)對車輛間肇事型態有顯著的影響關係。而中央分隔之形式於本研究中分為：

代號	情況
0	為單行道，並無中央分隔設置
1	以實體分隔之中央分隔島道路
2	雙向間雖無中央分隔島，但有反光標鈕分隔
3	雙向間僅以標線分隔

#### 5.快慢車道間分隔型式

同上，考量快車道與慢車道間之分隔形式對於安全的影響，本研究分類如下：

代號	情況
0	該行向僅有一混合車道，無快慢車道之分
1	快慢車道間以實體分隔
2	快慢車道間以反光標鈕分隔
3	快慢車道間僅以標線分隔

#### 6.左/右轉專用號誌

同有無設置左/右專行道，若一路口有大量的轉彎車輛，因於綠燈時，左轉車輛與對向直行車輛、右轉車輛與同向行人間均有衝突，故若設置左/右轉專用號誌，應可減少發生衝突與交通事故的機率。

#### 7.標誌個數

當駕駛人進入路口時，若其所面臨之交通標誌數量太多易使駕駛人分心，且過量之資訊亦容易導致混淆。

#### 8.時相複雜程度

一般路口為簡單二時相者，駕駛人常有不待自身行向之綠燈，而

見相鄰行向黃燈便搶先通行之現象。然而某些路口有左轉遲閉或單方向早開等特殊時相，容易造成衝突甚而引發肇事。因此時相之複雜程度亦列入考慮。以各行向所需面對之時相數計算之，而其中閃光號誌路口視為零時相。

#### 9.進入路口時是否有速限或警告標誌

若可於某些路口或T字路口設置清楚、明確之指示或警告標誌、標線及號誌，提醒駕駛人注意路口存在，進而達到提醒駕駛人依規定減速慢行與讓道，應可減少事故發生，故將此項因素列入考慮。

代號	情況
0	無有速限或警告標誌
1	有有速限或警告標誌

#### 10.車道被佔用程度

經路口實地調查時，發現市區道路常有路邊臨時違規停車，或道路兩旁民眾長期佔用，甚有位置不當之公共設施或停車位等等。這些狀況都會導致道路實際可使用之寬度大為減少，對於車流產生嚴重度不等的影響。雖然本研究之肇事限定於路口，然而由於臨近道路對路段若遭佔用，亦會影響車流。故將各行向臨近路口處道路被佔用之程度列為變數，放入模式中考量。研究將道路遭佔用分為四個等級：

代號	情況
0	道路兩旁未遭佔用，可完整使用慢車以及路肩
1	道路旁遭佔用，但僅及道路邊線以外部分
2	道路遭佔用，固定設施或常有違規停車佔用慢車道
3	道路遭佔用，固定設施或常有違規停車佔用快車道

#### 11.路旁十公尺內是否有障礙物

進行實地調查時，發現於停止線後十公尺路旁常有設置許多廣告與看板，這些設置不當之廣告與看板往往會影響駕駛人視野與注意



力，可能導致事故之發生。

代號	情況
0	路旁十公尺內無障礙物
1	路旁十公尺內有障礙物

#### 4.3.2 交通動態變數

在相同的幾何設計、時相配比、號誌標誌配置之環境下，因為不同的交通流量以及不同的組成，肇事之多寡依然會有所不同。因此除了工程因素變數外，希望在考量交通動態變數之因素下使預測更加準確。本研究所考量之交通動態變數分述如下：

##### 1. 交通量：

路口通過之車輛多，則彼此間之干擾與衝突亦隨之增多。因此進入路口的車流量可成為衡量肇事數量的影響因子，研究以交通量做為衡量肇事曝光量之基準。而在作為模式輸入時將路口之總交通量作為一輸入變數，除總交通量外亦將行向組合之各別交通量納入考量。

##### 2. 左/右轉車輛比

於路口進行轉向的車輛，對其他車輛形成干擾。其與同向之直行車輛形成匯出型衝突、與對向直行車輛形成交叉型衝突、與對向轉同一方向之車輛則形成匯入型衝突。因此路口轉彎車輛數之多寡對於衝突以及肇事的產生具有影響。

##### 3. 機車車輛比

不同車種所產生之肇事機率亦有所不同。研究將車流均換算為小客車當量數(Passenger Car Equivalent, PCU)，求出機車之比例，作為各行向之機車比之變數。

##### 4. 時制週期

若路口之時制週期過長，駕駛人常會不待本身之時向而違反交通

規則，而若路口常存有此種情況則將導致事故發生。因此將路口之時制週期納入模式考量因素之一。

### 5.綠燈秒數

在時制之綠燈秒數不符車流量所需，則駕駛人可能於綠燈結束與黃燈始亮時，強行穿越路口導致發生交通事故。

## 4.4 資料之蒐集及初步整理

本節將闡述相關肇事變數、工程因素變數及交通動態變數等資料之蒐集來源。分述如下：

### 4.4.1 路口及肇事相關資料之蒐集

過去研究中，楊憲國君以交通流量作為交通事故曝光量之基準，於 88、89 年擇選之台中市 62 個路口為研究對象；本研究後續選定台中市民國 92、93 年內曾進行過流量調查之 40 個路口，總計 102 個路口之交通事故資料為研究對象。經資料蒐集與整理後，將其拆解為行向組合之後，則可得到 1,039 組行向組合。此 102 個交會道路名稱如表 4.2 及 4.3 所示：

表 4.2 88、89 年選定路口之交會道路

路口編號	交會道路			路口編號	交會道路		
1	台中港路	東大路		12	台中港路	健行路	美村路
2	台中港路	工業一路		13	台中港路	英才路	
3	台中港路	安和路		14	台中港路	民權路	公益路
4	台中港路	光明路		15	台中港路	五權路	中正路
5	台中港路	朝富路		16	文心路	市政路	大墩十七街
6	台中港路	河南路		17	文心路	青海路	
7	台中港路	惠中路		18	文心路	河南路	
8	台中港路	文心路		19	文心路	大雅路	中清路
9	台中港路	大墩路	何厝路	20	文心路	崇德路	
10	台中港路	漢口路	東興路	21	忠明南路	國光路	
11	台中港路	忠明路	忠明南路	22	忠明南路	復興路	

表 4.2 88、89 年選定路口之交會道路 (續)

路口編號	交會道路			路口編號	交會道路		
23	忠明南路	三民西路		44	大雅路	英才路	
24	忠明南路	南屯路		45	大雅路	健行路	
25	忠明南路	五權西路		46	大雅路	太原路	
26	忠明南路	向上路		47	大雅路	忠明路	進化北路
27	忠明南路	公益路		48	大雅路	漢口路	
28	環中路	永春路		49	大雅路	天津路	
29	環中路	五權西路		50	大雅路	北平路	
30	環中路	市政路		51	公園路	三民路	
31	五權西路	忠勇路	安和路	52	精武路	雙十路	
32	五權南路	高工路		53	國光路	建成路	
33	五權南路	忠明南路		54	國光路	復興路	林森路
34	建成路	振興路	忠孝路	55	中清路	民航路	
	大興路			56	復興路	台中路	
35	建成路	旱溪街	東光園路	57	林森路	自由路	
36	崇德路	進化北路		58	林森路	五權路	英才路
37	崇德路	松竹路		59	公益路	英才路	
38	崇德路	昌平路	昌平十路	60	台中路	明德街	
39	崇德路	環中路		61	三民路	北屯路	雙十路
40	中清路	雷中街			健行路		
41	中清路	四平路		62	三民路	五權路	崇德路
42	中清路	水湳路	華美西街		錦南街		
43	大雅路	五權路	公園路				

表 4.3 本研究整理之路口

路口編號	交會道路			路口編號	交會道路		
1	台中港路	忠明南路		21	五權路	五權西路	
2	台中港路	文心路		22	文心路	青海路	
3	台中港路	河南路		23	文心路	河南路	
4	台中港路	惠來路		24	文心路	公益路	
5	五權南路	復興路		25	文心路	向上路	
6	五權南路	高工路		26	民權路	台中路	建國路
7	五權西路	美村路		27	民權路	自由路	
8	五權西路	環中路		28	民權路	三民路	
9	五權西路	忠勇路		29	民權路	五權路	
10	忠明南路	復興路		30	中正路	建國路	
11	忠明南路	三民西路		31	中正路	自由路	
12	忠明南路	南屯路		32	中正路	三民路	
13	忠明南路	五權西路		33	中正路	五權路	
14	忠明南路	向上路		34	三民路	自由路	南屯路
15	忠明南路	公益路		35	環中路	福科路	
16	忠明南路	五權南路		36	環中路	朝馬路	
17	文心路	市政路	大墩十七街	37	環中路	中清路	
18	文心南路	復興路		38	環中路	中清交流道	
19	文心南路	高工路		39	雙十路	自由路	
20	五權路	英才路	林森路	40	復興路	台中路	

#### 4.4.2 肇事資料之蒐集

為了將路口事故拆解成行向組合，從台中市警察局之資料庫發現並無對肇事車輛之行向有任何相關記錄，故在取得肇事資料之方式係經由台中市警察局交通隊之協助，以現場記錄與整理之方式登錄所需資料；基於考量保護肇事當事人資料，本研究所記錄之肇事資料僅含事故地點、時間、事故行向與結果。至於肇事經過摘要、當事人之姓

名、身份證字號、住址、電話及肇事車輛車號等則全未記錄。再此感謝台中市警察局交通隊提供之全力協助。

#### 4.4.3 交通動態資料之蒐集

因路口交通流量及轉向調查需要大量人力且成本高昂，因此本研究採用已有之資料，除原先楊憲國君於 88、89 年所甄選之路口外，並蒐集民國 92、93 年曾在台中市市區進行各路口流量及轉向調查之資料。因調查之時間不同，於各年間各路口之流量會有所變化，唯本研究採用車輛流量資料以做為衡量路口交通事故曝光量之衡量，故並不進行轉換計算之工作。

#### 4.3.4 道路環境資料之蒐集

本研究所需之路口資料，如道路幾何、標誌佈設及號誌時相等資料，均以實地進行測量與勘查。而各路口之時制相關資料如：綠燈秒數，則由台中市交通局交通工程科所提供，此一數值為為民國 92 年底之狀態。將所有取得之變數資料輸入資料庫中，以供模式使用。

上述各點所處理的流程如下圖所示：

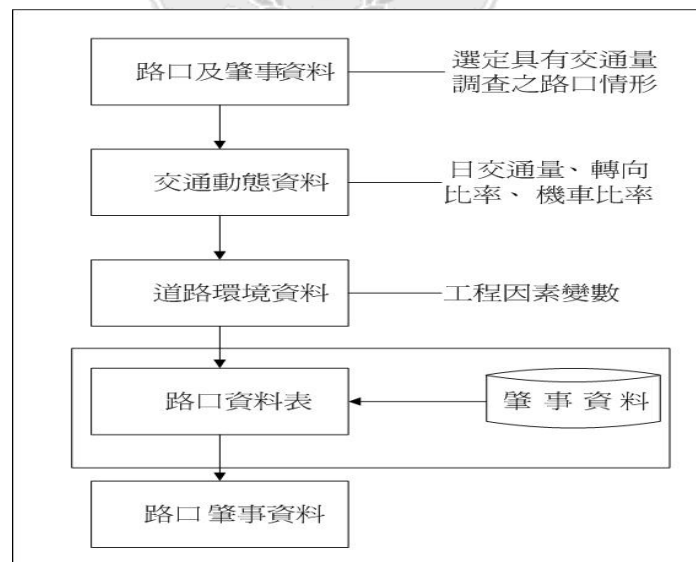


圖 4.2 資料整理流程圖

#### 4.5 路口交通事故資料初步分析

本研究先以事故資料作初步分析，針對本研究整理之台中市路口

事故之次數、行向組合之次數及路口肇事時段之分析。分述如下：

#### 4.5.1 路口事故之次數

經資料蒐集整理，本研究所採取之資料為民國 88、89 年間所選定之 62 個路口及民國 92、93 年間所選定之 40 個路口，其各時間所發生之交通事故為 1,593 件及 1,848 件，總計 3,441 件交通事故（如表 4.4 及 4.5）。而圖 4.2 及圖 4.3 中，民國 88、89 年發生之路口交通事故多半集中於 30 次以下，其中以 10~20 次最多；而事故次數超過 30 次以上雖僅佔少數，但仍有超過事故次數為 100 次以上之路口。相較民國 92、93 年，所發生之次數則有增無減，多數集中在 10~50 次間，以 20~30 最多，甚至有路口事故次數高達近 200 次，次數佔調查期間約 10.4%，顯得相當具危險。

表 4.4 88、89 年路口之交通事故次數

路口編號	事故次數	路口編號	事故次數	路口編號	事故次數	路口編號	事故次數
1	15	17	14	33	24	49	21
2	9	18	12	34	78	50	8
3	21	19	15	35	29	51	22
4	14	20	30	36	31	52	20
5	20	21	39	37	52	53	26
6	14	22	13	38	29	54	19
7	27	23	20	39	50	55	17
8	10	24	25	40	10	56	24
9	18	25	33	41	9	57	28
10	21	26	21	42	11	58	10
11	9	27	21	43	11	59	11
12	5	28	133	44	14	60	82
13	10	29	36	45	12	61	27
14	8	30	31	46	16	62	26
15	15	31	68	47	18		
16	12	32	62	48	14		

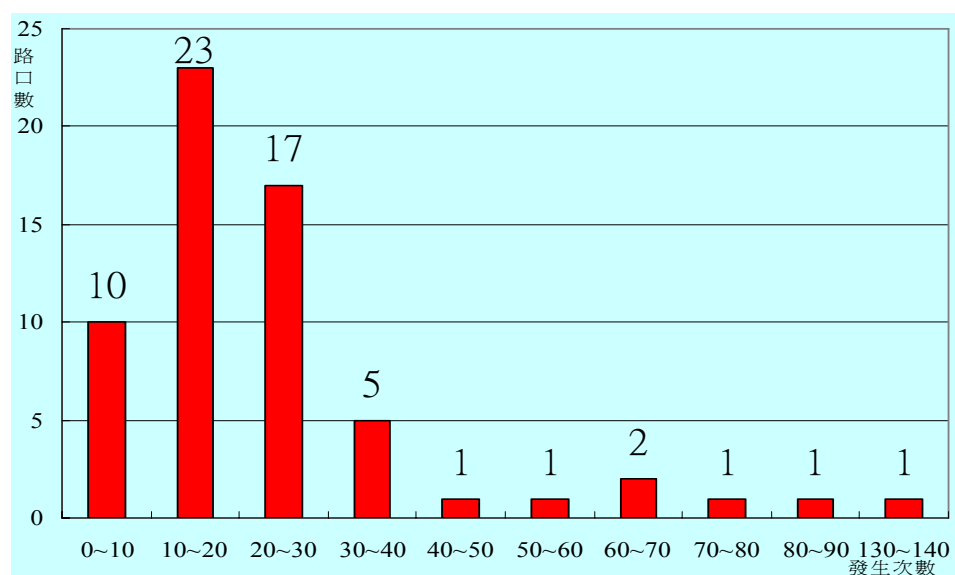


圖 4.3 88、89 年路口之交通事故次數分佈圖

表 4.5 92、93 年之交通事故次數

路口編號	事故次數	路口編號	事故次數	路口編號	事故次數	路口編號	事故次數
1	88	11	14	21	24	31	20
2	194	12	6	22	41	32	20
3	105	13	35	23	41	33	45
4	76	14	22	24	45	34	17
5	54	15	27	25	38	35	65
6	47	16	23	26	42	36	49
7	28	17	55	27	20	37	123
8	137	18	44	28	24	38	22
9	23	19	28	29	35	39	31
10	24	20	30	30	45	40	40

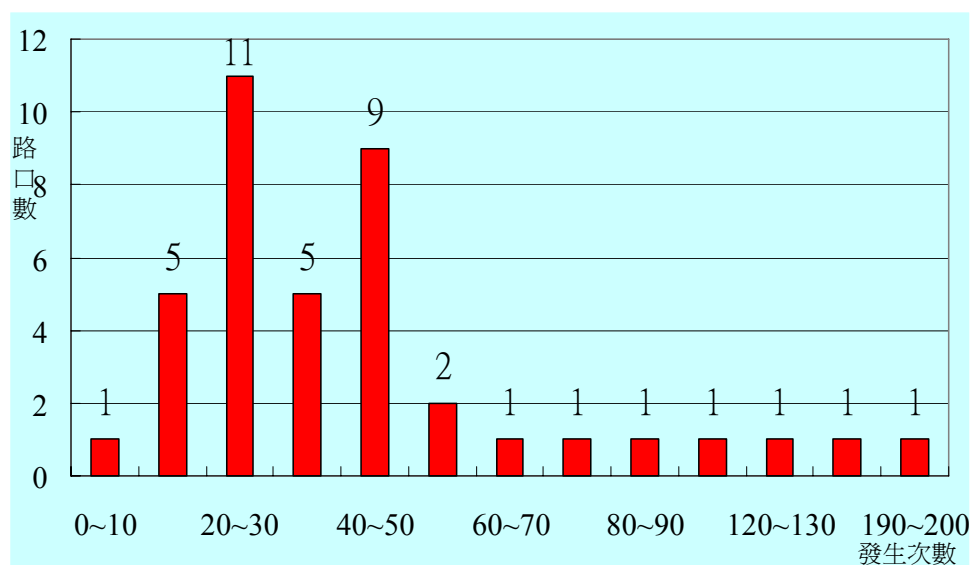


圖 4.4 92、93 年路口之交通事故次數分佈圖

#### 4.5.2 相異時間之路口事故次數

由路口事故之次數中，發現本研究選取之路口，部份與民國 88、89 年有重複選取，為了精確探求路口彼此處在相異空間造成之事故次數，對此 18 個路口(表 4.6)進行分析。發現絕多數路口之事故發生皆有明顯上升，其中以台中港路與文心路路口最多，五權西路與環中路路口次之；相對亦有路口之事故次數降低，如忠明南路與南屯路路口。

表 4.6 相異時間路口總事故次數彙整表

路口編號	路口名稱	民國 88、89 年事故次數	民國 92、93 年事故次數
1	台中港路與河南路	78	105
2	台中港路與文心路	133	194
3	台中港路與忠明南路	62	88
4	中正路與五權路	21	45
5	五權路與英才路	14	30
6	文心路與市政路	39	55
7	文心路與青海路	20	41
8	文心路與河南路	13	41
9	忠明南路與復興路	18	24



表 4.6 相異時間路口總事故次數彙整表(續)

路口編號	路口名稱	民國 88、89 年事故次數	民國 92、93 年事故次數
10	忠明南路與三民西路	9	14
11	忠明南路與南屯路	12	6
12	忠明南路與五權西路	11	35
13	忠明南路與向上路	14	22
14	忠明南路與公益路	11	27
15	五權西路與環中路	82	137
16	五權西路與忠勇路	15	23
17	五權南路與高工路	14	47
18	復興路與台中路	10	40

為了瞭解肇事次數遞增或遞減之原因，將以交通事故發生次數最多之台中港路與文心路路口及肇事次數降低之忠明南路與南屯路路口進行分析。首先，台中港路與文心路路口之基本調查資料與路口示意圖如表 4.7 與圖 4.5 所示：

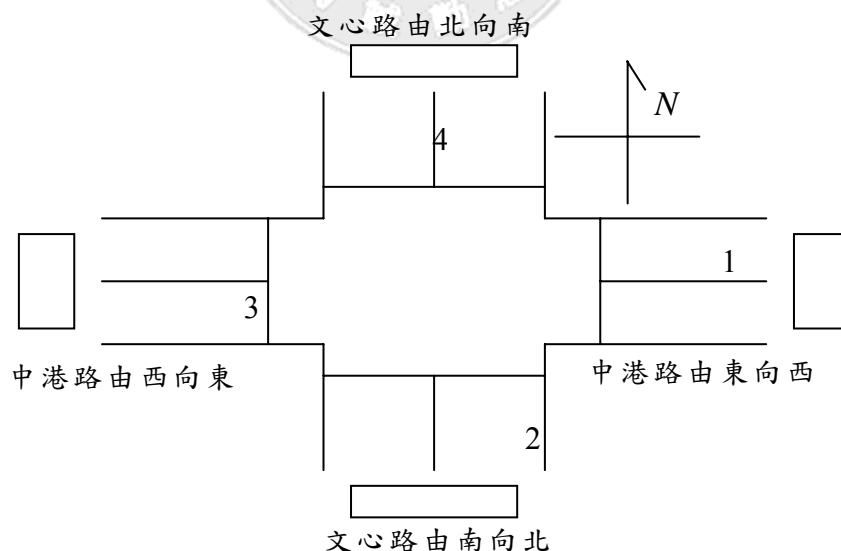


圖 4.5 台中港與文心路交叉路口示意圖

表 4.7 台中港路與文心路路口之基本調查資料

時間	民國 88、89 年				民國 92、93 年			
行向代碼	1	2	3	4	1	2	3	4
道路名稱	台中港路	文心路	台中港路	文心路	台中港路	文心路	台中港路	文心路
車輛行向	東	南	西	北	東	南	西	北
快車道數	4	4	4	4	4	4	4	4
中央分隔	實體分隔	實體分隔	實體分隔	實體分隔	實體分隔	實體分隔	實體分隔	實體分隔
快慢分隔	實體分隔	標線	實體分隔	標線	實體分隔	標線	實體分隔	標線
左轉專用道	2	1	2	1	2	1	2	1
道路寬度(m)	23.6	18	24.3	16.9	23.6	18	24.3	16.9
道路被佔用程度	未遭佔用	未遭佔用	未遭佔用	未遭佔用	未遭佔用	未遭佔用	未遭佔用	未遭佔用
時相複雜程度	5	4	6	4	5	5	5	5
號誌個數(個)	5	5	5	5	5	5	5	5
左轉用號誌	有	有	有	有	有	有	有	有
號誌周期	180	180	180	180	180	180	180	180
綠燈秒數(秒)	60	30	60	30	44	40	44	40
行向流量(pcu)	7,341	4,410	8,938	5,363	6,492	3,901	7,065	4,181
機車比	0.247	0.418	0.321	0.383	0.313	0.439	0.370	0.493
左轉比	0.111	0.234	0.241	0.338	0.111	0.179	0.239	0.391
右轉比	0.134	0.099	0.089	0.114	0.096	0.077	0.058	0.098
限速或警告標誌	有	有	有	有	有	有	有	有
十公尺內是否有障礙物	是	否	否	否	是	否	否	否

台中港路與文心路路口為資料中肇事次數最多，亦是肇事次數上升最高之肇事地點，實際事故次數分別為 133 件及 194 件。就其行向組合事故而言，民國 88、89 年間，以行向 1 與行向 3 之組合其發生次數最多，其次為行向 1 與行向 1、行向 3 與行向 3 之組合；民國 92、93 年間卻發現行向為同向之組合發生較多次數，以行向 4 與行向 4 之組合最多(如表 4.8)。

由表 4.7 及表 4.8 中發現，短短 3 年間，路口之道路幾何設計皆維持原先之規劃，如中央分隔帶設置、快慢車分隔帶設置、左轉專用道及道路寬度等等；反觀，道路動態則有差異，故可推斷肇事次數增加，交通動態與道路幾何設計間彼此具有相互關聯性。以行向 4 與行向 4 之組合為例，道路寬度僅 16.9 公尺，為本路口路寬最小，而行向流量減少，機車比率之增加，且左轉車輛比增加、快慢車分隔帶設置以標線劃分，故若駕駛者因路口流量減少導致速度過快，與近 5 成之機車比率及左專車輛比率近 4 成匯集在路口，其衝突點繁多，因而較易發生事故。

表 4.8 中港路與文心路路口各行向組合之交通事故實際次數

行向組合	事故次數		行向組合	事故次數	
	88、89 年	92、93 年		88、89 年	92、93 年
1&1	28	15	2&3	5	9
1&2	8	4	2&4	2	12
1&3	38	8	3&3	28	28
1&4	5	5	3&4	4	4
2&2	5	53	4&4	10	56

忠明南路與南屯路路口之基本調查資料與路口示意圖如表 4.9 與圖 4.6 所示：

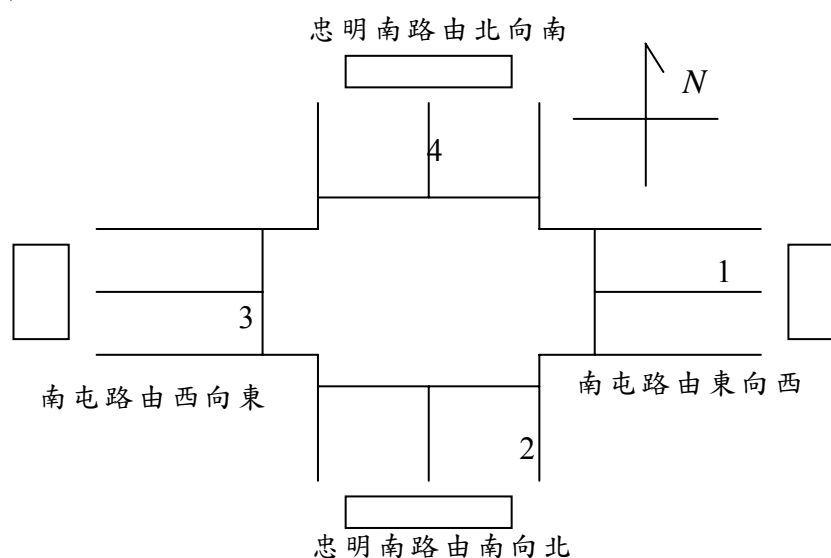


圖 4.6 忠明南路與南屯路交叉路口示意圖

表 4.9 忠明南路與南屯路路口之基本調查資

時間	民國 88、89 年				民國 92、93 年			
行向代碼	1	2	3	4	1	2	3	4
道路名稱	南屯路	忠明南路	南屯路	忠明南路	南屯路	忠明南路	南屯路	忠明南路
車輛行向	東	南	西	北	東	南	西	北
快車道數	2	3	2	3	2	3	2	3
中央分隔	標線	實體分隔	標線	實體分隔	標線	實體分隔	標線	實體分隔
快慢分隔	標線	標線	標線	標線	標線	標線	標線	標線
左轉專用道	1	1	1	1	1	1	1	1
道路寬度(m)	8.6	12.8	9.5	12.6	8.6	12.8	9.5	12.6
道路被佔用程度	佔用慢車道	未遭佔用	佔用慢車道	未遭佔用	未遭佔用	未遭佔用	未遭佔用	佔用慢車道
時相複雜程度	4	4	4	4	4	4	4	4
號誌個數(個)	5	5	5	5	5	5	5	5
左轉用號誌	有	有	有	有	有	有	有	有
號誌周期	150	150	150	150	150	150	150	150
綠燈秒數(秒)	35	40	35	40	35	40	35	40
行向流量(pcu)	1,773	2,642	2,544	1,172	1,354	2,877	2,376	2,070
機車比	0.509	0.397	0.442	0.329	0.576	0.458	0.558	0.503
左轉比	0.043	0.462	0.203	0.058	0.025	0.250	0.233	0.086
右轉比	0.100	0.001	0.086	0.015	0.079	0.031	0.174	0.092
限速或警告標誌	無	無	無	無	無	無	無	無
十公尺內是否有障礙物	是	是	是	是	是	否	否	否

忠明南路與南屯路路口於此空間之肇事次數並未如其他路口增加，反是降低 6 件。就其行向組合事故而言，民國 88、89 年間，以行向 1 與行向 1、行向 1 與行向 2 及行向 4 與行向 4 之組合為無事故次數發生；92、93 年間以行向 1 與行向 1、行向 1 與行向 2、行向 1 與行向 3、行向 1 與行向 4、行向 2 與行向 3 及行向 3 與行向 4 之組合為無事故次數發生(如表 4.10)。

由表 4.9 與表 4.10 得知，忠明南路與南屯路路口，僅有道路動態資料變異，雖差異於交通動態，但路口事故之形成絕非憑一論點即可果斷。以行向 3 與行向 3 之組合為例，民國 88、89 年路口上游十尺內有障礙物佔據慢車道，極可能影響慢車道駕駛者之行車視距，導致事故發生；而民國 92、93 年間，即使機車比率及右轉車輛比率增加，但因路口十尺內不具任何障礙物，故未發生任何事故。

表 4.10 忠明南路與南屯路路口各行向組合之交通事故實際次數

行向組合	事故次數		行向組合	事故次數	
	88、89 年	92、93 年		88、89 年	92、93 年
1&1	0	0	2&3	1	0
1&2	1	0	2&4	3	1
1&3	0	0	3&3	4	1
1&4	1	0	3&4	1	0
2&2	1	3	4&4	0	1

#### 4.5.3 行向組合交通事故次數

經台中市警察局交通隊之協助，本研究依據其肇事現場圖判定肇事車輛之行向，將 3,441 件事故案例拆解，即得 1,039 組行向組合之事故發生次數。本研究即以此 1,039 筆資料，做為建構模式之用。此 1,039 組行向組合經整理所蒐集之交通事故資料後，呈現出偏左之行向組合肇事次數分佈圖(圖 4.7)。圖中得知事故次數分佈普遍偏低，其中以發生次數為一次及零次之行向組合總數最多，分別佔 23.58%及 22.33%，其後為隨之遞減。發生次數為 12 次以上，所佔比率皆不達 1%，部分發生次數甚至存在著行向組合總數為 1 組，在差異如此極大之比例，是故將造成本研究建構模式之預測能力降低。

表 4.11 行向組合交通事故次數統計表

發生次數	行向組合總數	所佔比例(%)	發生次數	行向組合總數	所佔比例(%)
0	232	22.33	15	6	0.58
1	245	23.58	16	3	0.29
2	159	15.30	17	3	0.29
3	127	12.22	19	1	0.10
4	60	5.77	20	5	0.48
5	39	3.75	21	3	0.29
6	32	3.08	22	2	0.19
7	30	2.89	23	1	0.10
8	21	2.02	25	3	0.29
9	21	2.02	28	5	0.48
10	11	1.06	30	2	0.19
11	11	1.06	38	1	0.10
12	2	0.19	53	1	0.10
13	7	0.67	56	1	0.10
14	5	0.48	共計	1,039	100.00

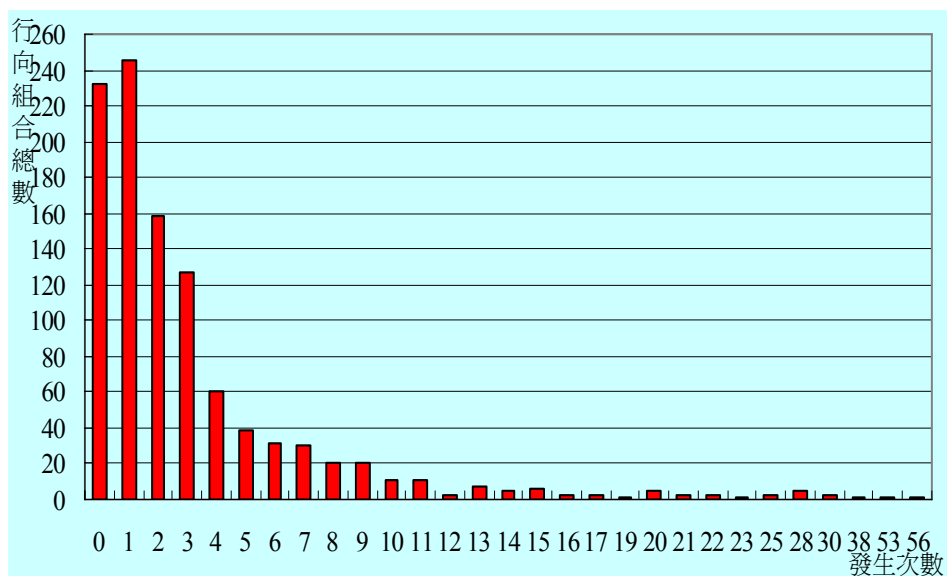


圖 4.7 行向組合肇事次數分佈圖

#### 4.5.4 路口肇事時段

而在肇事時間的分佈上，選定路口之肇事時間分佈如表 4.12 及圖 4.8 所示。

表 4.12 選取路口肇事時段統計表

時段	肇事次數	所佔百分比	時段	肇事次數	所佔百分比
0—1	132	3.84%	12—13	152	4.42%
1—2	80	2.32%	13—14	147	4.27%
2—3	59	1.71%	14—15	153	4.45%
3—4	72	2.09%	15—16	178	5.17%
4—5	56	1.63%	16—17	186	5.41%
5—6	92	2.67%	17—18	199	5.78%
6—7	100	2.91%	18—19	222	6.45%
7—8	214	6.22%	19—20	146	4.24%
8—9	195	5.67%	20—21	121	3.52%
9—10	154	4.48%	21—22	172	5.00%
10—11	161	4.68%	22—23	171	4.97%
11—12	138	4.01%	23—24	141	4.10%
總計		3,441	100.00%		

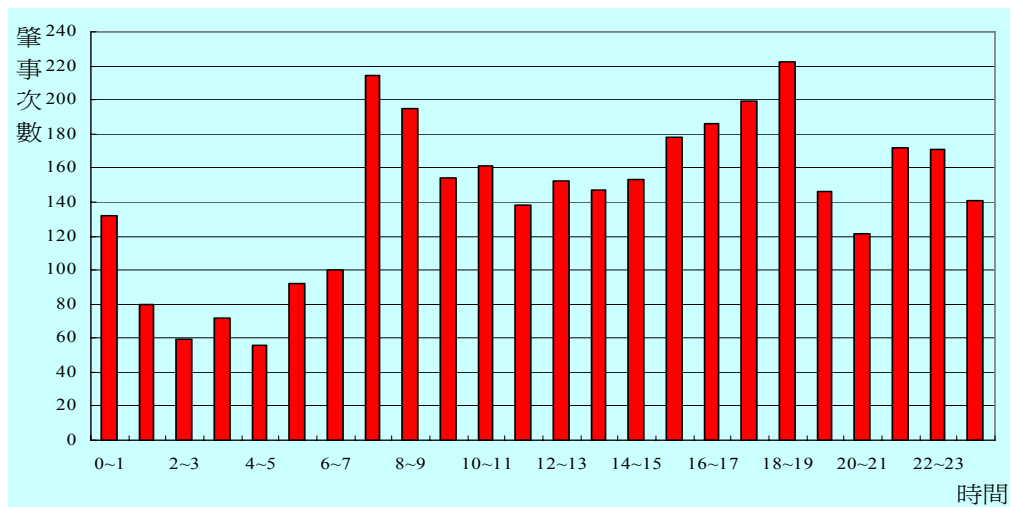


圖 4.8 選取路口肇事時段分佈圖

對於路口肇事時段分佈中，7 時至 9 時及 16 時至 19 時具有高發生頻率之趨勢，其次為 21 時至 23 時，於 1 時至 5 時發生頻率低。其中以 7~8 時、17~18 時及 18~19 時為發生頻率最高的三個時段，而於此三個時段為上下班時間之尖峰時間。其原因可能為於尖峰時段交通流量較大、車流密度較高，車輛發生碰撞的機率亦隨之增加。而於凌晨及午夜時段 1 時至 5 時發生頻率最低。





## 第五章 模式建構與分析

本章將彙整完成之路口肇事資料以統計方法與類神經網路模式建構與分析，並且將其結果做比較，分述如下：

### 5.1 敘述性統計分析

本研究所採用之相關變數如第四章所述，將整理出的路口肇事資料透過敘述性統計分析，其彙整如表 5.1 所示：

表 5.1 變數資料之基本統計

變數名稱	平均值	標準差	最大值	最小值
肇事次數	3.22	4.94	56	0
快車道個數	2.74	1.13	4	2
左轉專用道	0.75	0.62	2	0
道路寬度	13.45	6.92	58.9	1.6
中央分隔設施型式	1.74	0.99	3	0
快慢車道間分隔型式	2.63	0.86	3	0
號誌個數	4.24	0.91	5	0
左轉專用號誌	0.56	0.50	1	0
時相複雜程度	3.64	0.71	7	2
時制週期	155.79	18.67	180	90
綠燈秒數	50.12	20.74	115	0
速限警告標誌	0.17	0.38	2	0
路旁十公尺內是否有障礙物	0.42	0.49	1	0
車道佔用程度	0.44	0.54	3	0
路口總流量	12,027.87	3,895.14	26,052	5,448
行向流量	2,918.50	1,693.84	9,025	0
左轉比率	0.20	0.18	0.93	0
右轉比率	0.16	0.16	1	0
機車比率	0.42	0.14	0.96	0

表 5.1 中各路口行向之年肇事次數平均在 3.22 次，其中最大次數為 56 次，表示所觀測路口行向每年事故最多發生 56 次；各車道之左轉專用道，最多為 2 個，平均為 0.75 個；路口總流量在各路口中最大為 26,052 輛次，最小為 5,448 輛次，平均為 12,027.87 輛次；各行向流量平均約 2,919 輛次；平均左/右轉車輛比各為 0.2/0.16 輛次，平均機車比為 0.42 輛次。從本研究所採用之變數發現，可知各變數之值域範圍差異甚大，如路口總交通流量之值域範圍介於  $10^3$  與  $10^4$  之間；行向流量之值域範圍介於 0 與  $10^3$  之間；而機車比率、轉彎比率之值域範圍均介於 0 與 1 之間。

## 5.2 統計模式與分析

從文獻中可知，統計迴歸技術於交通事故因果模式之建構以具廣泛被應用，本研究亦嘗試統計方法來分析路口肇事次數與其相關之肇事因子，於建立迴歸技術之前，將其檢定迴歸模式較符合卜瓦松迴歸模式 (Poisson Regression Model) 或是負二項迴歸模式 (Negative Binomial Regression Model)，再以該模式建立與分析。由於檢定結果適合負二項迴歸模式，所以以此模式來說明建立與分析之結果。

### 5.2.1 負二項模式建立

由於肇事次數屬於計數性資料，過去常使用的統計模式為卜瓦松迴歸或是負二項迴歸，若資料的期望值與變異數不相等時，可透過負二項迴歸中  $\alpha$  參數是否顯著來判斷資料適合的模式。

過去路口或路段之事故相關研究，許多學者為了使其預測之準確度提高，而提出許多假設，如楊憲國君【21】因行向組合發生事故次數為零次之資料相對於其他次數發生筆數高，認為在進行負二項迴歸模式建構時，容易造成模式偏估，無法建立適用模式，故研究將肇事次數為零之資料去除；Shankar【50】認為發生頻率記錄為零時，不必然代表該路段較安全，其可能較安全而未發生事故或曾發生事故卻並未記錄，亦可能調查時間並未發生事故，而假設迴歸模式將事故次數為零的情況設定一機率值。

茲因若未考量發生事故為零次之行向組合，則對於路口之幾何設計與交通動態之安全範圍無依參考。故本研究先行納入事故頻率為零次，其進行負二項迴歸模式建構時，發現相關係數(R)值為 0.498，誤判率高達 87.23%，故參照 Shankar 之想法，設定事故為零次值為 0.001，再進行模式建構。

建構模式時，將行向組合資料筆數以隨機之方式抽取 75%(即 779 筆)為訓練筆數。再以其餘 25%(即 260 筆)的行向組合資料予以驗證。所輸入的目標變數(Target Variable)為各路口行向每年肇事次數，而預測變數(Predictors)共輸入 18 個變數，其中除時制週期與路口總流量外，其餘預測變數皆需考慮兩車行進方向，共計 34 個變數，使用軟體為 Limdep 7.0 的 Negative Binomial Regression Model 進行參數校估與檢定。所建立之負二項迴歸模式校估與檢定結果如表 5.2 所示。

表 5.2 變數資料之基本統計

變數名稱	代號	係數	P value	變數名稱	代號	係數	P value
常數項		-0.7399	0.1534	速限或警告標誌 2	【 X <sub>21</sub> 】	-0.1767	0.1395
快車道數 1	【 X <sub>1</sub> 】	-0.05314	0.1054	十公尺內有障礙 1	【 X <sub>22</sub> 】	-0.1899	0.0741
快車道數 2	【 X <sub>2</sub> 】	-0.03471	0.1139	十公尺內有障礙 2	【 X <sub>23</sub> 】	-0.0505	0.6272
左轉專用道 1	【 X <sub>3</sub> 】	-0.0006	0.9949	車道佔用程度 1	【 X <sub>24</sub> 】	0.0145	0.8762
左轉專用道 2	【 X <sub>4</sub> 】	0.1658	0.1135	車道佔用程度 2	【 X <sub>25</sub> 】	-0.2357	0.0089
道路寬度 1	【 X <sub>5</sub> 】	0.0169	0.0141	路口總交通量	【 X <sub>26</sub> 】	0.000043	0.0011
道路寬度 2	【 X <sub>6</sub> 】	-0.0046	0.0603	交通量 1	【 X <sub>27</sub> 】	0.0001	0.0624
中央分隔設施型式 1	【 X <sub>7</sub> 】	-0.0123	0.8262	交通量 2	【 X <sub>28</sub> 】	0.0001	0.0151
中央分隔設施型式 2	【 X <sub>8</sub> 】	0.0350	0.5477	左轉車輛比 1	【 X <sub>29</sub> 】	0.1256	0.6168
快慢車道分隔型式 1	【 X <sub>9</sub> 】	0.0151	0.7915	左轉車輛比 2	【 X <sub>30</sub> 】	0.1182	0.6466
快慢車道分隔型式 2	【 X <sub>10</sub> 】	-0.0824	0.1956	右轉車輛比 1	【 X <sub>31</sub> 】	0.4424	0.1151
號誌個數 1	【 X <sub>11</sub> 】	0.1960	0.0226	右轉車輛比 2	【 X <sub>32</sub> 】	-0.1679	0.5338
號誌個數 2	【 X <sub>12</sub> 】	-0.1861	0.0256	機車比 1	【 X <sub>33</sub> 】	0.0689	0.0671
左轉用號誌 1	【 X <sub>13</sub> 】	0.0687	0.7007	機車比 2	【 X <sub>34</sub> 】	0.0758	0.0481
左轉用號誌 2	【 X <sub>14</sub> 】	0.2787	0.1253	$\alpha$ (alpha)		0.7252	0.0000
時相複雜程度 1	【 X <sub>15</sub> 】	-0.1186	0.0145	Degrees of freedom		32	
時相複雜程度 2	【 X <sub>16</sub> 】	0.0489	0.0756				
時制週期	【 X <sub>17</sub> 】	0.0053	0.0313				
綠燈秒數 1	【 X <sub>18</sub> 】	0.0009	0.7657	Log likelihood function		-1465.274	
綠燈秒數 2	【 X <sub>19</sub> 】	0.0004	0.8964	Restricted log likelihood		-1851.917	
速限或警告標誌 1	【 X <sub>20</sub> 】	0.3470	0.0044	Chi-squared ( $\chi^2$ )		43.2857	

建立之路口交通事故分析模式如式 5.1 所示：

$$u = \exp(-0.7399 - 0.05314 X_1 - 0.03471 X_2 - 0.0006 X_3 + 0.1658 X_4 + 0.0169 X_5 \\ - 0.0046 X_6 - 0.0123 X_7 + 0.0350 X_8 + 0.0151 X_9 - 0.0824 X_{10} + 0.1960 X_{11} \\ - 0.1861 X_{12} + 0.0687 X_{13} + 0.2787 X_{14} - 0.1186 X_{15} + 0.0489 X_{16} + 0.0053 X_{17} \\ + 0.0009 X_{18} + 0.0004 X_{19} + 0.3470 X_{20} - 0.1767 X_{21} - 0.1899 X_{22} - 0.0505 X_{23} \\ + 0.0145 X_{24} - 0.2357 X_{25} - 0.000043 X_{26} + 0.0001 X_{27} + 0.0001 X_{28} + 0.1256 X_{29} \\ + 0.1182 X_{30} + 0.4424 X_{31} - 0.1679 X_{32} + 0.0689 X_{33} + 0.0758 X_{34}) \dots\dots\dots (5.1)$$

其中，

u：為交通事故次數

X<sub>1</sub>：為行向組合中其一行向之快車道數

X<sub>2</sub>：為行向組合中另一行向之快車道數

X<sub>3</sub>：為行向組合中其一行向之左轉專用道

X<sub>4</sub>：為行向組合中另一行向之左轉專用道

X<sub>5</sub>：為行向組合中其一行向之道路寬度

X<sub>6</sub>：為行向組合中另一行向之道路寬度

X<sub>7</sub>：為行向組合中其一行向之中央分隔設施型式

X<sub>8</sub>：為行向組合中另一行向之中央分隔設施型式

X<sub>9</sub>：為行向組合中其一行向之快慢車道分隔型式

X<sub>10</sub>：為行向組合中另一行向之快慢車道分隔型式

X<sub>11</sub>：為行向組合中其一行向之號誌個數

X<sub>12</sub>：為行向組合中另一行向之號誌個數

X<sub>13</sub>：為行向組合中另一行向之左轉用號誌

X<sub>14</sub>：為行向組合中另一行向之左轉用號誌

X<sub>15</sub>：為行向組合中另一行向之時相複雜程度

X<sub>16</sub>：為行向組合中另一行向之時相複雜程度

X<sub>17</sub>：為行向組合中之時制週期

X<sub>18</sub>：為行向組合中其一行向之綠燈秒數

X<sub>19</sub>：為行向組合中其一行向之綠燈秒數

X<sub>20</sub>：為行向組合中其一行向之速限或警告標誌

X<sub>21</sub>：為行向組合中其一行向之速限或警告標誌

X<sub>22</sub>：為行向組合中其一行向之十公尺內有障礙

$X_{23}$ ：為行向組合中其一行向之十公尺內有障礙

$X_{24}$ ：為行向組合中其一行向之車道佔用程度

$X_{25}$ ：為行向組合中其一行向之車道佔用程度

$X_{26}$ ：為路口之總交通量

$X_{27}$ ：為行向組合中另一行向之交通量

$X_{28}$ ：為行向組合中另一行向之交通量

$X_{29}$ ：為行向組合中其一行向之左轉車輛比

$X_{30}$ ：為行向組合中另一行向之左轉車輛比

$X_{31}$ ：為行向組合中其一行向之右轉車輛比

$X_{32}$ ：為行向組合中另一行向之右轉車輛比

$X_{33}$ ：為行向組合中另一行向之機車比

$X_{34}$ ：為行向組合中另一行向之機車比

從模式建立的結果發現， $X^2$  值為  $43.2857 < \chi^2_{0.05}$  臨界值為 46.1701，檢定過度離散參數  $\alpha$  是否顯著來作離散性檢定，其結果發現， $\alpha$  值為 0.7252，過適參數  $\alpha$  在 95% 顯著水準下已達顯著 ( $P$  值 =  $0.0000 < 0.05$ )，顯示負二項迴歸模式較為適合。模式中有 9 個變數在  $\alpha = 0.05$  水準下達顯著，分別為道路寬度、號誌個數、時相複雜程度、時制週期、速限或警告標誌、車道佔用程度、路口總交通量、行向流量與機車比率；另外一個變數在  $\alpha = 0.1$  以下達顯著，為十公尺內有障礙。

經由模式解釋變數係數之正負符號來解釋對於肇事次數發生機率的增減，本研究為尋求預測結果，則將測試資料代入迴歸式，反推回肇事次數的機率值，並以最大的機率值為預測的肇事次數。其預測驗證結果如圖 5.2 所示，圖中橫軸(T)代表的是實際值，縱軸(A)代表的是模式預測值，最佳結果為預測值等於實際值，即為通過原點之虛線。實線為代表整體驗證資料之線段。

經其預測結果與實際值比較，發現模式於預測高交通事故次數之行向組合中，預測值具有低估之趨勢；交通事故次數較低的行向組合相較於高交通事故次數之行向組合而言，高、低估情況皆有，相關係數值為 0.843。就模式預測值與實際值之差值，如圖 5.1 所示，若預測

值與實際值相差於 0.5 以內則視為判中，其餘則視為誤判，其誤判率高達 57.42%，其差值最大為-25.43。

由圖中可得知於高次數之行向組合中，以負二項迴歸模式預測值較實際值低估的情況較多，此種情況之產生可能為所蒐集之行向組合之交通事故次數大多數為低次數居多，資料具有之偏態分佈之情況，無法代表高交通事故次數之行向組合之特性，而導致建構之模式於預測高次數之行向組合時，產生低估之情況。

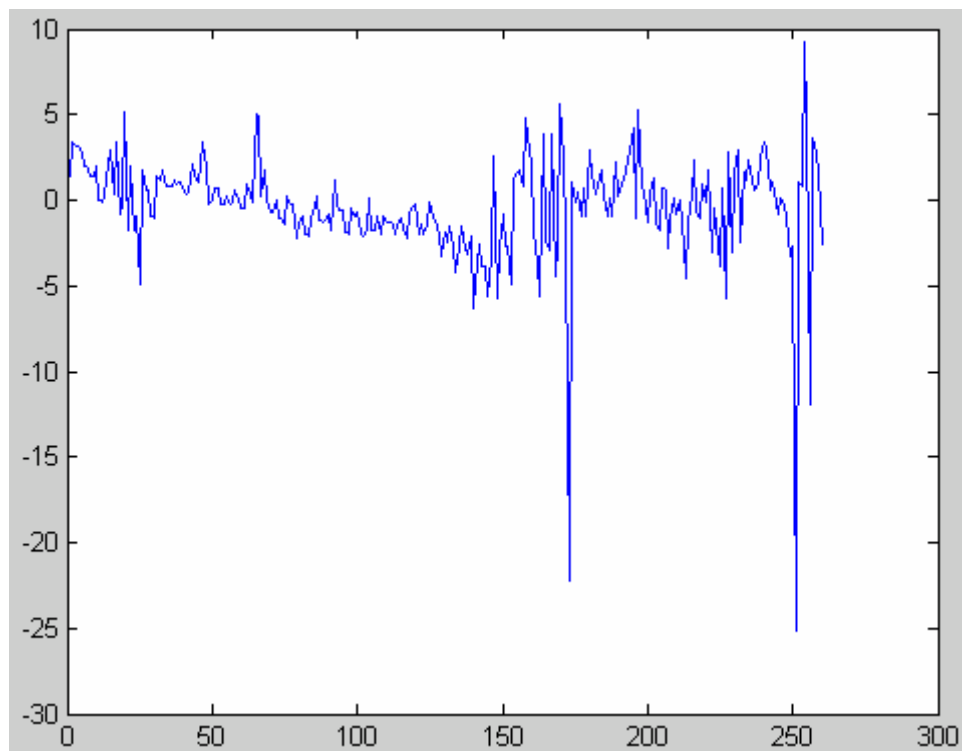


圖 5.1 負二項迴歸模式預測值與實際值之差值圖

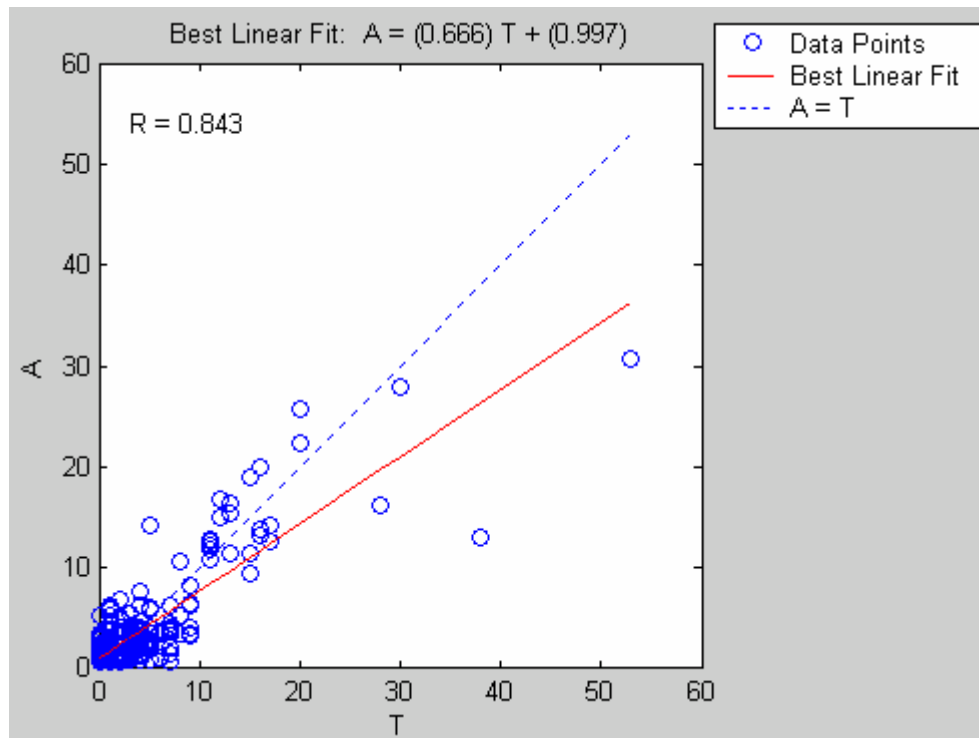


圖 5.2 負二項迴歸模式預測圖

### 5.2.2 彈性分析

對所建構之負二項迴歸模式進行彈性分析，以評估模式中獨立變數的變化對於相依變數之影響。以本研究而言，就是獨立變數變動情形對於事故次數之影響。茲因本研究選取之變數涵蓋類別變數及連續變數兩部份，是故在求彈性變數時，需使用適合之公式，如連續變數之彈性，稱為直接彈性(Direct Elasticity)，計算方式為

$$E_{X_{ij}}^{\lambda_i} = \beta_j X_{ij} \dots\dots\dots (5.2)$$

其中E表示彈性值， $\beta_j$ 為 $X_{ij}$ 之係數， $X_{ij}$ 為觀測值i其變數j之值， $\lambda_i$ 為觀測值i其期望次數，所以 $\lambda_i$ 為路段i之平均事故次數， $X_{ij}$ 可為路段i之交通量。

當變數為虛擬變數(Indicator Variable)時或是該變數之值只有0或1時，則要透過擬似彈性(Pseudo-Elasticity)來計算，計算公式為

$$E_{X_{ij}}^{\lambda_i} = \frac{\exp(\beta_j) - 1}{\exp(\beta_j)} \dots\dots\dots (5.3)$$

經彈性分析之結果如表5.3所示，因彈性值之正負符號可解釋對於肇事次數發生機率之增減，故本研究採取其絕對值依序由大至小，則為道路寬度、路口總交通量、機車比率、時相複雜程度、時制週期及行向流量等等。顯示交通動態變數遠比道路幾何設計規劃之變數，對於事故之影響絕佔大數。如為低流量、時相複雜性小且道路寬度較寬之路口，即使行駛速率不會過快，假若通過路口之機車數量過多，此時機車駕駛產生不當駕駛行為，易導致事故之發生。

表 5.3 負二項迴歸模式變數彈性值

變數名稱	彈性值	變數名稱	彈性值	變數名稱	彈性值
快車道個數 1	-0.03	左轉用號誌 1	0.01	車道佔用程度 2	-0.03
快車道個數 2	-0.01	左轉用號誌 2	0.05	路口總交通量	0.15
左轉專用道 1	0.02	時相複雜程度 1	-0.13	行向流量 1	0.07
左轉專用道 2	0.04	時相複雜程度 2	0.06	行向流量 2	0.09
道路寬度 1	0.17	時制週期	0.11	左轉車輛比 1	0.01
道路寬度 2	-0.07	綠燈秒數 1	0.01	左轉車輛比 2	0.01
中央分隔設施型式 1	-0.01	綠燈秒數 2	0.01	右轉車輛比 1	0.02
中央分隔設施型式 2	0.02	速限或警告標誌 1	0.02	右轉車輛比 2	-0.01
快慢車道分隔型式 1	0.01	速限或警告標誌 2	-0.01	機車比 1	0.14
快慢車道分隔型式 2	-0.07	十公尺內有障礙 1	-0.02	機車比 2	0.09
號誌個數 1	0.06	十公尺內有障礙 2	-0.01		
號誌個數 2	-0.05	車道佔用程度 1	0.05		

## 5.3 類神經網路模式與分析

### 5.3.1 類神經網路模式建立

本研究擬採用目前應用相當普遍的倒傳遞類神經網路(Back-Propagation Network, BPN)模式。為充分捕捉變數間複雜的互動現象，以監督式學習模式為主。所謂監督式學習方法是有正確的目標輸出值(Target Output)作為依據，以多組訓練範本(Training Patterns)含輸入及目標輸出值來訓練與測試。無回饋且完全連結(Full-Connected)



之向前網路(Feed-Forward Network)，來建構路口肇事次數預測模式，網路依輸入值計算網路輸出值，並和該輸入的目標輸出值比較，依此調整連接權重，重複訓練至誤差在容許範圍內。期望能藉其加入隱藏層並使用非線性作用函數(Non-Line Active Function)之特質，有效反應危險影響因子之交互作用並學習非線性的資料型態，進而達到正確判斷危險程度之目的。

雖類神經網路之輸入值可包含任意大小之變數值，但仍於不同變數值之間範圍差別太大，可能會影響網路之學習效率。因此於實際操作上，可將各輸入處理單元之值域轉換至 $[-1,1]$ 或 $[0,1]$ 間，可得到較佳之結果。從表 5.1 發現，各變數之值域範圍差異甚大，為了減少資料變異程度，提升預測準確率，於隱藏層轉換函數以正切雙彎曲轉換函數將資料正規化(Normalizing)，將其值域轉換至 $[-1,1]$ 之間。其公式如下：

$$f(x) = \frac{e^n - e^{-n}}{e^n + e^{-n}} \dots\dots\dots (5.4)$$

其中，

$f(x)$  為轉換後之值

$n$  為輸入之自變數值

本研究在進行模式之建構與驗證前，先將總資料 1,039 筆肇事資料分為模式訓練組與模式驗證組。故在總資料中將抽取約 75%(即 779 筆)做為訓練類神經網路之資料筆數，以開始進行模式訓練以建立其預測模式。再以其餘 25%(即 260 筆)的行向組合資料對於所建立之模式予以驗證。利用軟體 MATLAB 6.5 所提供之環境自行撰寫相關模式所需程式，進行建構與驗證。

為了得到模式最佳預測之績效，訓練類神經網路之前，首先要進行網路訓練模式的設定，各設定如表 5.4 所示，本研究嘗試不同學習演算法並針對隱藏層神經元數目做調整。其中輸入層與隱藏層之轉換函數採用正切雙彎曲轉換函數(Sigmoid Threshold Transfer Function)，隱藏層與輸出層之間之轉移函數為線性轉換函數(Linear Transfer Function)。

表 5.4 網路相關設定值

網 路 相 關 參 數 名 稱	設 定 值
隱藏層數量	1 層
隱藏層神經元數目	10 個、15 個、20 個、25 個、30 個
隱藏層轉換函數	正切雙彎曲轉換函數
輸出轉換函數	線性轉換函數
輸出單元數	1 個
學習演算法	共軛梯度演算法 擬牛頓演算法
設定訓練次數	3500 次
設定性能目標	$10^{-6}$

將本研究 1,039 筆事故資料中抽取 75% 先進行各模式之訓練與學習，25% 資料對於所建立之模式予以驗證。各模式網路預測值與實際值比較結果與相關係數值，如表 5.5 所示。

模式訓練驗證後得知預測結果與實際結果之相關係數是以演算法為共軛梯度演算法且隱藏層神經元 25 個之模式所得 0.967 最大，判中率達 73.33%，而訓練結果之判中率是以 20 個隱藏層神經元之模式所得 100% 最高。演算法為共軛梯度演算法時，其結果看來輸入模式之隱藏層神經元數在超過 20 個之後，在 20 個與 30 個神經元時其相關係數及判中率呈現僵固的狀態，就驗證績效的觀點來看，25 個隱藏層神經元即可達到最佳預測結果。

而擬牛頓演算法，雖隱藏層神經元數目增加，其相關係數及判中率仍持續微緩上升，不過就驗證績效的觀點來看，隱藏層神經元數目高達 30 節點數時，都尚未能突破六成之判中率。因此，本研究學習演算法以共軛梯度演算法為設定，再細部針對隱藏層神經元數於 20 個至 25 個，試求找出最佳化模型，結果於隱藏層神經元數為 21 個，此時網路之 MSE 值為  $2.98 \times 10^{-6}$  驗證相關係數值為 0.994，判中率為 86.47%，為本研究之最佳化模式，其相關結果如表 5.6。

表 5.5 相異演算法及隱藏層節點數之訓練驗證結果

演算法	隱藏層神經元數目	訓練組		測試組	
		相關係數	判中率	相關係數	判中率
共軛梯度演算法	10	0.866	61.27%	0.703	47.49%
	15	0.931	67.33%	0.834	56.25%
	20	1	100%	0.954	71.81%
	25	1	100%	0.967	73.33%
	30	1	100%	0.955	72.45%
擬牛頓演算法	10	0.746	53.16%	0.691	34.63%
	15	0.856	60.84%	0.733	50.34%
	20	0.937	68.23%	0.782	53.41%
	25	0.975	72.14%	0.836	56.82%
	30	0.985	73.37%	0.884	58.63%

表 5.6 相異隱藏層節點數之訓練驗證結果

隱藏層神經元數目	訓練組		測試組	
	相關係數	判中率	相關係數	判中率
20	1	100.00%	0.954	71.81%
21	1	100.00%	0.994	86.47%
22	1	100.00%	0.986	81.12%
23	1	100.00%	0.975	74.06%
24	1	100.00%	0.969	73.35%
25	1	100.00%	0.967	73.33%

若將建構網路之資料以本研究所建構之網路模式進行模擬，其資料筆數為 779 筆，網路預測值與實際值比較結果與相關係數值，如圖 5.4 所示，圖中橫軸(T)代表的是實際值，縱軸(A)代表的是模式預測值，最佳結果為預測值等於實際值，即為通過原點之虛線，實線為代表整體資料趨勢之線段。而代表最佳結果之虛線與代表訓練資料整體趨勢之實線，產生重疊之情況，相關係數值為 1，以及由圖 5.3 中，將模式結果四捨五入，預測值與實際值相差於 0.5 以內則視為判中，然其餘則視為誤判，其預測模式預測值與實際值完全相同，表示類神經網路模式對於模擬訓練資料而言，其預測能力相當準確。

將驗證資料以建構之類神經網路模式，進行測試其預測值與實際值比較結果與相關係數值，如圖 5.6 所示，代表最佳結果之虛線與代表訓練資料整體趨勢之實線，幾乎產生重疊之情況，相關係數值為 0.994，而由圖 5.5 中模式預測值與實際值之差值來說，其誤判比率為 13.53%，其差值皆於  $\pm 1.5$  之間，顯示所建構之類神經網路模式雖仍有判預測值與實際值之差距仍屬可接受之範圍。

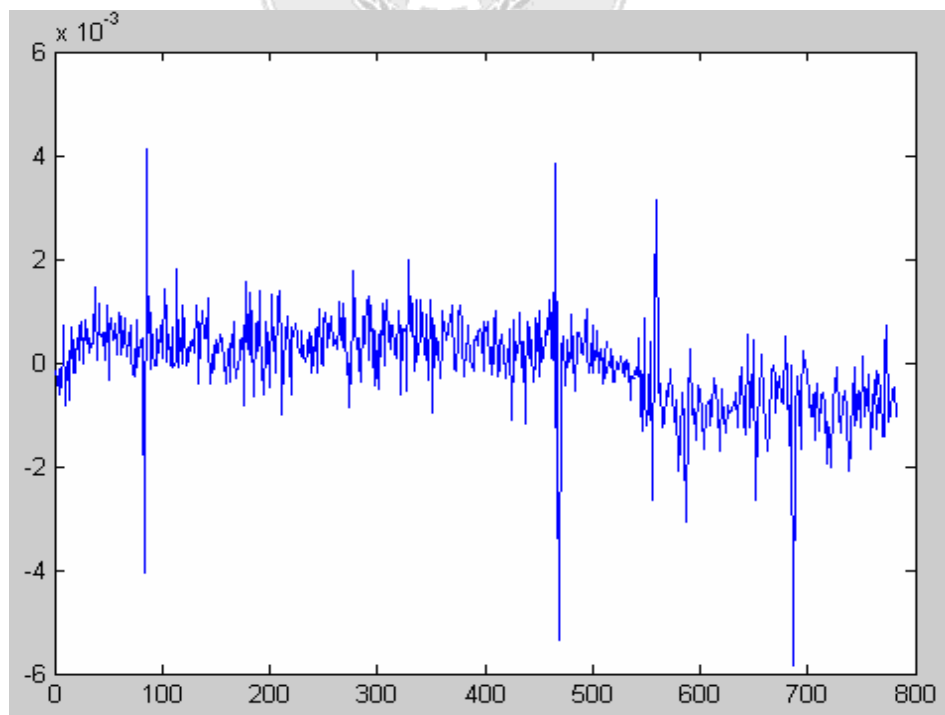


圖 5.3 類神經網路模式模擬訓練資料預測值與實際值之差值圖

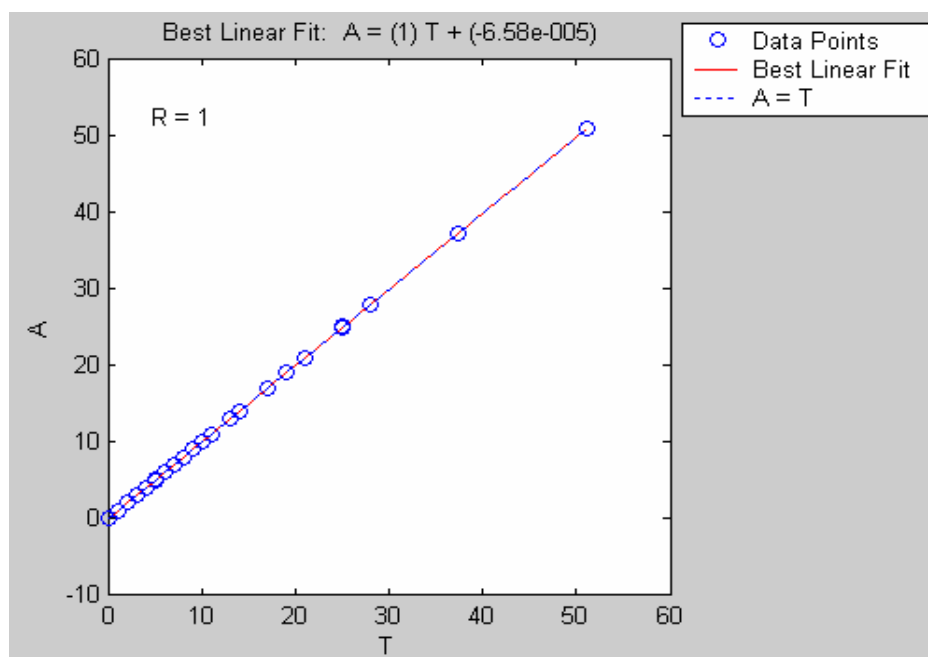


圖 5.4 類神經網路模式模擬訓練資料預測圖

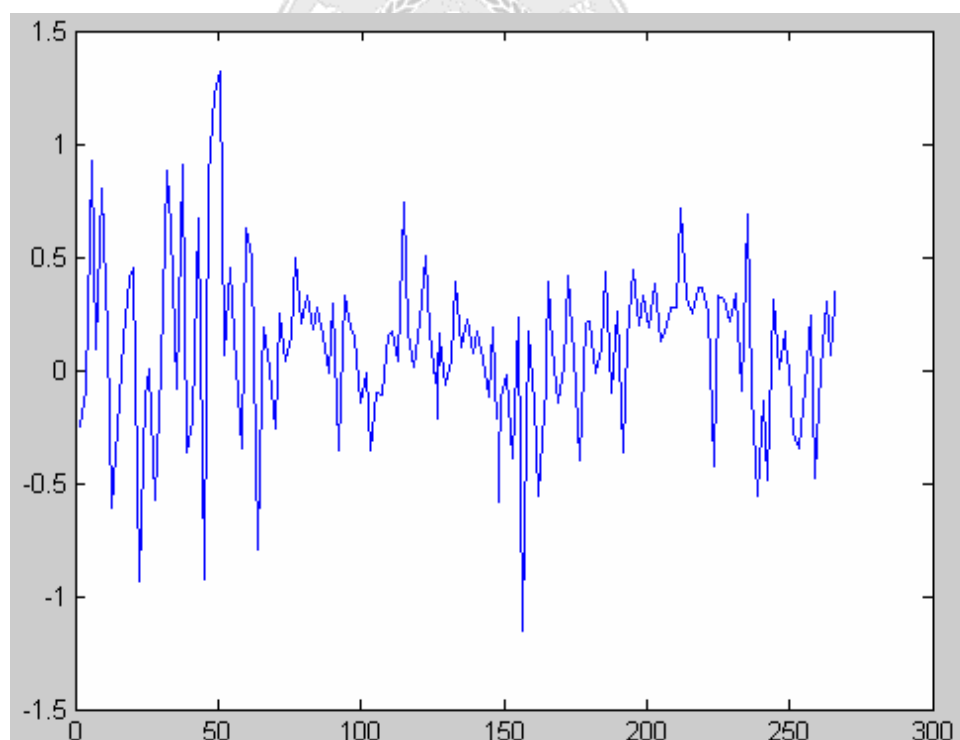


圖 5.5 類神網路模式模擬驗試資料預測值與實際值之差值圖

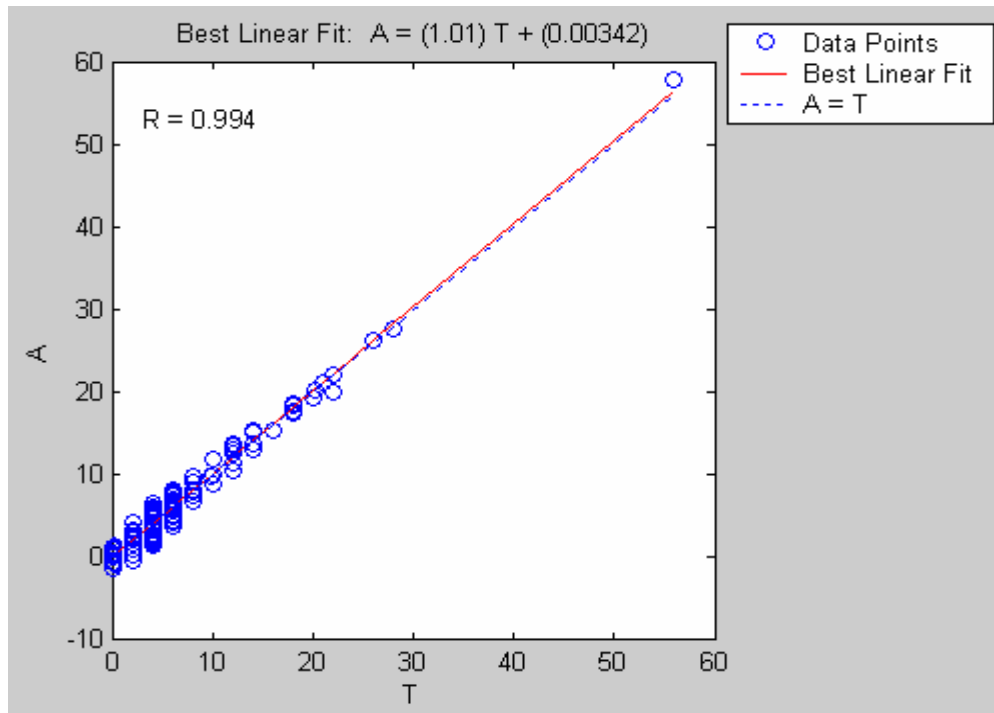


圖 5.6 類神經網路測試資料預測圖

### 5.3.2 敏感度分析

類神經網路模式是以權重型態記錄在網路中，但若以此複雜之架構記錄之權重而言，實在無法讓使用者瞭解其中所隱含的意義。因此，則有一些學者從事類神經網路擷取機制之研究，其主要之目的是將權重矩陣中之知識庫轉化成法則知識，讓使用者可以瞭解分析模式中所隱含的意義，進而分析其中因子之關係。

因此，若利用類神經網路的模擬並加上敏感度分析，確定每個變數在網路模式中之影響程度，估計每個輸入變數之影響值，即可將造成路口肇事次數之誤差降低，提高網路學習之效能。

經由類神經網路權重矩陣中擷取各因子之敏感度，其計算如公式所示。以敏感度擷取模式擷取肇事次數因子之敏感度為例，計算過程如下：

設 $X_i$ 為輸入層第 $i$ 個處理單元輸入值

$H_k$ 為隱藏層第 $k$ 個處理單元之輸出值

$Y_j$ 為輸出層第 $j$ 個處理單元之推論輸出值

$net_j$ 為輸出層第 $j$ 個處理單元之集成函數

$net_k$ 為隱藏層第 $k$ 個處理單元之集成函數。

則 $X_i$ 對 $Y_j$ 透過隱藏層第 $k$ 個處理單元之敏感度 $(S_{ij})^k$ 如下所示：

$$(S_{ij})^k = \left[ \frac{\partial Y_j}{\partial X_i} \right]^k = \left[ \frac{\partial Y_j}{\partial net_j} \right] \cdot \left[ \frac{\partial net_j}{\partial H_k} \right] \cdot \left[ \frac{\partial H_k}{\partial net_k} \right] \cdot \left[ \frac{\partial net_k}{\partial X_i} \right] \dots\dots\dots (5.5)$$

$$= f'(net_j) \cdot W_{ij} \cdot f'(net_k) \cdot W_{ik}$$

若假設 $f'(net_j)$ 與 $f'(net_k)$ 為常數，忽略不計，則可得敏感度 $S_{ij}$ ，如下述公式：

$$S_{ij} = \sum_k (S_{ij})^k = \sum_k W_{kj} \cdot W_{ik} \dots\dots\dots (5.6)$$

其中

$W_{kj}$ 為第 $k$ 個隱藏層單元與第 $j$ 個輸出層單元間之連結權重

$W_{ik}$ 為第 $i$ 個輸入層單元與第 $k$ 個隱藏層單元間之連結權重

當網路訓練完畢後，將此網路中各節點的權值以及門檻值，利用式5.5與5.6便可求得各輸入變數相對輸出變數的敏感度值。

各事故影響因子之敏感度分析結果如表 5.7 所示。

表 5.7 類神經網路輸入變數之敏感度平均值

變數名稱	敏感值	變數名稱	敏感值	變數名稱	敏感值
快車道個數 1	0.02	左轉專用號誌 1	0.06	車道佔用程度 2	0.03
快車道個數 2	-0.01	左轉專用號誌 2	0.04	路口總交通量	-0.11
左轉專用道 1	-0.04	時相複雜程度 1	0.04	行向流量 1	0.05
左轉專用道 2	0.07	時相複雜程度 2	-0.12	行向流量 2	0.04
道路寬度 1	0.25	時制週期	0.12	左轉車輛比 1	-0.05
道路寬度 2	0.00	綠燈秒數 1	0.07	左轉車輛比 2	0.35
中央分隔設施型式 1	0.05	綠燈秒數 2	-0.01	右轉車輛比 1	-0.02
中央分隔設施型式 2	0.01	速限或警告標誌 1	0.18	右轉車輛比 2	0.02
快慢車道分隔型式 1	0.01	速限或警告標誌 2	0.07	機車比 1	-0.22
快慢車道分隔型式 2	-0.08	十公尺內有障礙 1	-0.04	機車比 2	0.05
號誌個數 1	0.00	十公尺內有障礙 2	0.05		
號誌個數 2	0.09	車道佔用程度 1	0.03		

表 5.7 中的敏感度代表各個因子對於是否發生事故之影響程度，正值代表正相關，負值代表負相關，而數字的大小代表該因子之影響程度。部分因子之敏感度呈現負值，表示對其整個網路模組之影響為負相關，此結果可能與一般所理解之情形有所出入；其主要原因為類神經網路是針對整體因子作非線性之分析，各因子間相互之影響可能造成某一因子之敏感度完全壓過受其關聯之因子，使該關聯因子之敏感度無法完全顯現出來，因而造成負相關之情形產生。

分析結果顯示，以輸入變數「左轉專用道」為例，其敏感度值如圖 5.7 所示，其敏感度分析結果並非一定值，大多數之敏感度值集中於 0 附近，其中有數筆資料之敏感值明顯不為零。其可能原因為每組行向組合其構成因子相異，同一因子變化對於不同行向組合之影響亦有所不同，例如：當增加左轉專用道數時，對不同之路口幾何環境、號誌設施、與交通流量特向而言，對交通事故之發生次數可能增加，亦可能減少。說明如後：對於左轉車輛比率較大且流量較大之路口，其增加左轉專用道數有助紓解交通壅塞，減少車輛攔長，降低交通事故之發生次數；而對於左轉車輛比率較小且流量較大之路口之路口，其增加左轉專用道數因快車道數不足，增加車輛間碰撞機會，提高事故發生次數。故於不同之路口幾何、號誌及交通現況因素而言，單一變數的改變對於交通事故次數之影響並非一定值。

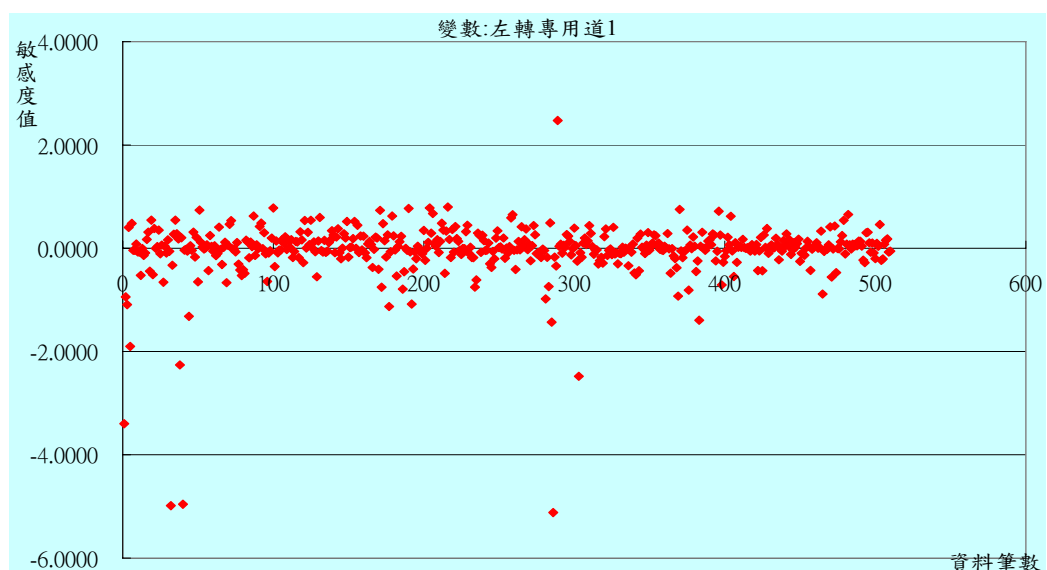


圖 5.7 左轉專用道之敏感度分佈圖



本研究欲找出影響路口交通事故之重要因素，因對於不同之行向組合其分析值可能為正值亦可能為負值，為防止正值與負值相消取其值後導致其值趨近於零，故對於模式之變數其敏感度分析先取絕對值再平均之結果如表 5.8 所示，其中以變數之敏感分析值由大至小為道路寬度、行向流量、左轉車輛比、機車比、時相複雜程度、道路總交通量及右轉車輛比等等。

表 5.8 敏感度值先取絕對值後再平均之結果

變數名稱	敏感值	變數名稱	敏感值	變數名稱	敏感值
快車道個數 1	0.18	左轉專用號誌 1	0.28	車道佔用程度 2	0.25
快車道個數 2	0.16	左轉專用號誌 2	0.28	路口總交通量	0.34
左轉專用道 1	0.26	時相複雜程度 1	0.38	行向流量 1	0.43
左轉專用道 2	0.27	時相複雜程度 2	0.36	行向流量 2	0.29
道路寬度 1	0.49	時制週期	0.27	左轉車輛比 1	0.44
道路寬度 2	0.28	綠燈秒數 1	0.19	左轉車輛比 2	0.26
中央分隔設施型式 1	0.24	綠燈秒數 2	0.28	右轉車輛比 1	0.32
中央分隔設施型式 2	0.13	速限或警告標誌 1	0.27	右轉車輛比 2	0.29
快慢車道分隔型式 1	0.21	速限或警告標誌 2	0.25	機車比 1	0.34
快慢車道分隔型式 2	0.22	十公尺內有障礙 1	0.15	機車比 2	0.32
號誌個數 1	0.22	十公尺內有障礙 2	0.25		
號誌個數 2	0.29	車道佔用程度 1	0.26		

## 5.4 負二項模式與類神經網路模式比較

本節以負二項模式與類神經網路模式做比較。

### 5.4.1 預測正確率比較

評估準確性的標準一般有誤差均方根(Root Mean Square, RMS)、平均絕對誤差(Mean Absolute Deviation, MAD)、相關係數(Correlation Coefficient)以及誤判率(Negligence Rate)等。因本研究為預測事故次數，若模式之預測值接近實際值或與實際值相同，模式之相關係數值越高與誤判率越低，則所建構之模式其預測能力越佳、適用性越高。

是故本研究選取相關係數與誤判率來進行模式間之比較分析。比較結果如表 5.9 所示。

表 5.9 類神經網路與負二項迴歸模式之比較

項目	相關係數	誤判率(%)	最大誤差
類神經網路模式	0.994	13.53%	1.37
負二項迴歸模式	0.843	57.42%	-25.43

以相關係數而言，類神經網路之相關係數為 0.994 比負二項迴歸模式之 0.843 較佳。就誤判率而言，類神經網路之誤判率為 13.53%，相對於負二項迴歸模式 57.42%之高誤判率，類神經網路之誤判率較低；而就其模式預測產生之最大誤差而言，亦為類神經網路較佳。經由上述之比較若就模式之預測能力而言，以類神經網路之預測能力較佳。

#### 5.4.2 敏感度分析與彈性之比較

敏感度分析與彈性分析結果，變數重要性排序整理如表 5.10 得知，各處理分析所得到之重要項目不盡相同，其中出現頻率最高為道路寬度、機車比、時相複雜程度、路口總交通量，其次分別為左轉機車比、時制週期及行向流量，最後才為速限或警告標誌、快慢車分隔設施及右轉車輛比。

表 5.10 敏感度與彈性分析之比較

項目 排序	整體資料之敏感度分析	取絕對值後整體資料之敏感度分析	彈性分析
1	左轉車輛比	道路寬度	道路寬度
2	道路寬度	行向流量	路口總交通量
3	機車比	左轉車輛比	機車比
4	速限或警告標誌	機車比	時相複雜程
5	時相複雜程度	時相複雜程度	時制週期
6	時制週期	道路總交通量	行向流量
7	路口總交通量	右轉車輛比	快慢車分隔形式

## 5.5 小結

交通事故之發生具有偶然性與不可預測性，其影響因素眾多且彼此間相互影響，本研究僅考量路口幾何、誌號與交通現況等因子作為模式之輸入變數，其他影響因子並未納入模式中。而於輸入變數方面，是利用現地調查與事故發生當時之路口幾何、號誌及交通現狀因素，亦有所不同。負二項迴歸模式假設事件發生的期望計數值符合一個無法觀察到的伽瑪分配(Gamma Distribution)，而所蒐集之資料是否符合此分配並無法得知，這些因素均可能導致變數校估結果不顯著與負二項迴歸模式之適用性不佳。

由敏感度分析與彈性分析評選出之重要因素，各處理分析結果所得之重要項目不盡相同，其原因可能為行向組合其構成因子相異，同一因子變化對於不同行向組合之影響亦有所不同。而為了進一步瞭解路口幾何、號誌及交通動態與交通事故數量間之相互關係，將歸納所得之 10 個變數，採用資料探勘之分類與迴歸樹，挖掘路口幾何、號誌及交通動態與交通事故數量間之規則，並找出左轉車輛比、道路寬度、機車比、行向流量、速限或警告標誌、時相複雜程度、時制週期、路口總交通量、右轉車輛比及快慢車分隔形式在路口危險程度中所扮演之角色。

## 第六章 道路危險性指標與決策樹分析

為了提昇駕駛者行車安全，降低易肇事地點之危險程度，政府每年均投注大量人力與經費於易肇事地點的改善，但現行作業對於易肇事路口主要肇因之判定、改善措施之研擬及改善效果之評量，並無一套完整之作業規範、準則或實證結果。有鑑於此，本研究乃以台中市之 102 個路口為研究對象，藉由分類與迴歸樹分析，輔助改善計畫中對路口主要肇因之判定。

### 6.1 道路潛在危險性評估指標

道路潛在危險性評估指標係在反應道路交通事故「多次」發生之地點中(可為路段、路口或特定地點)所隱藏之潛在危險。因此，透過道路潛在危險性評估指標之審慎設計，可用來評估道路所具有之潛在危險性，進而據此找出高風險之地點，以供改善。

本研究認為一個合適之道路潛在危險性評估指標，必須能讓分析人員利用此指標辨識出須改進之分析對象時，即可運用工程技術來降低具高危險性對象，其危險性應是肇因於「路」的因素多過肇因於「人」或「車」之因素。此外本研究亦認為一個好的指標，對例行性與專案性質之評估工作，且應具有足夠之適用性，如此方能確實滿足分析人員之需求。

為建立上述一套肇事地點危險程度之判斷法則，以供實務界依此法則而施行，必須選擇一個良好之道路潛在危險性評估指標。本研究依據運輸研究所於民國 86 年之「道路潛在危險性評估指標之研究」中，發現目前國內外對於辨識危險對象之技術探討已久，亦發展出多種不同之方法論，整體而言大概可歸為肇事次數法、肇事率法、肇事機率法、交通阻礙法、事故嚴重性法其他多種方法，其各方法的優異特性簡述如表 6.1。

表 6.1 國內外辨識危險對象之主要方法論彙整表

評估方法	方法摘要	特性	優點	缺點
肇事次數法	統以發生肇事次數之多寡來判定分析對象的潛在危險性。	<ul style="list-style-type: none"> <li>◆較適於分析小城市之街道系統。</li> <li>◆亦適於分析大城市交通流量低的道路系統。</li> <li>◆不適於分析交通流量高或變化大的道路系統。</li> </ul>	<ul style="list-style-type: none"> <li>◆方法簡單易懂。</li> <li>◆資料可直接、輕易地取得。</li> </ul>	<ul style="list-style-type: none"> <li>◆忽略曝光量因素之影響，使比較之基礎不一。</li> <li>◆未考量事故之傷亡情形。</li> </ul>
肇事率法	將肇事次數除以曝光量，即得肇事率，以此值之高低來判定分析對象的潛在危險性。	<ul style="list-style-type: none"> <li>◆改善照是次數法在比較基礎上不平等的缺失。</li> </ul>	<ul style="list-style-type: none"> <li>◆可消除曝光量因素對肇事次數高低之影響。</li> <li>◆可使比較立基於較一致的水平上。</li> <li>◆方法易懂。</li> </ul>	<ul style="list-style-type: none"> <li>◆會因為曝光量而產生肇事率偏高或偏低的現象。</li> <li>◆未考量事故之傷亡情形。</li> </ul>
肇事機率法	應用品管理論，為各分析對象建立其在特定信賴水準下的肇事率上下限。	<ul style="list-style-type: none"> <li>◆可因應各分析對象之曝光量情形，變動其上下限區域。</li> </ul>	<ul style="list-style-type: none"> <li>◆可使比較立基於較一致的水平上。</li> <li>◆可適度描繪實際值與統計值間差異的隨機性。</li> </ul>	<ul style="list-style-type: none"> <li>◆方法較複雜。</li> <li>◆評估者須對統計理論有所瞭解。</li> <li>◆須假設道路交通事故資料服從某些分配。</li> </ul>

表 6.1 國內外辨識危險對象之主要方法論彙整表(續)

評估方法	方法摘要	特性	優點	缺點
交通阻礙法	為各種肇事結果、「交通特性」、「人」及「路」因素建立指示值，並加權計算一分析對象之各項考慮因素的指示值，以求得其潛在危險性指標。	◆非如其他方法僅單純採用肇事結果(肇事次數、肇事率、嚴重程度)資訊，來判斷一分析對象之潛在危險性。	◆兼顧交通、環境與其他構成一分析對象潛在危險性之因素。	◆分析所需之資料甚難蒐集。 ◆權數的訂定見仁見智。 ◆各因素指示值的訂定，難有客觀標準。
事故嚴重性法	利用死亡、受傷及財損成本資料或直觀論定，產生權數來加權一分析對象之死亡、受傷與僅有財損之肇事次數，得出以僅有財損為基本當量之肇事次數當量值，最後以此值來評估分析對象的潛在危險性。	◆較適於交通事故之經濟分析。 ◆通常會對有人死亡之事故給予最高之權重，其次為有人受傷之事故，而僅有財損之事故的權數為一(基本當量)。此案是分析對象之事故結果愈嚴重時，其危險性越高。	◆只要權數一經決定，則評估方法甚為簡單。 ◆兼顧肇事次數與傷亡情形。 ◆方法易懂。	◆權數的決定，難有客觀定論。 ◆當一分析對象發生稀少的嚴重事故時，此方法會凸顯其危險性，可能肇政工程改善不具成本效益。 ◆未考量曝光因素時的影響。

由前述諸多方法中可發現，找出具潛在危險性之分析對象的方法甚多，但絕大多數之方法運用幾種常用的基本統計資料，如：肇事次數(道路交通事故發生的次數)、死亡人數、受傷人數與肇事率(特定曝光量下道路交通事故發生的次數)，透過加權、排序、統計理論與多重參照等方式，找出一套較適用的篩選邏輯，有助於決定較合適之危險對象辨識之方法。

茲因本研究所蒐集之資料僅針對道路環境因子與肇事次數間之影響程度，故以利用發生道路交通事故頻率多寡來判定危險性高低之作法，除資料易取得外，對於道路交通事故發生次數多的對象而言，此法將具相當之危險性，而此種危險性的成因，依此法的判斷邏輯，係肇因於道路因素。因為分析對象與眾多不同之人、車而言，都會發生道路交通事故，即表示其危險性源自於非人、車因素，而所謂非人車因素，即為道路因素；對於低肇事次數但傷亡情形嚴重的對象而言，此法將可能排除危險性，進而忽略了此種分析對象之道路因素所可能造成之危險性。

## 6.2 模式建構

由諸多方法中，本研究採以「肇事次數法」作為判斷危險之準則。而辨識路口危險程度實務上並無類似之判斷準則，故本研究嘗試將蒐集之行向組合資料，依其交通事故次數由小至大排序，取其第 75 百分位數(Third Quartile)之事故次數值 3，與同資料平均交通事故次數 3.22 相近，故以此作為判別行向組合危險性之臨界值，給予語意「普通危險」。應因肇事次數為 1 次與未發生事故之路口行向佔全部觀測數目的比例最高，其次為 2~4 次，而 5 次以上之肇事次數佔全部觀測數的比例最低。

為降低在 CART 分析中之複雜度，將此輸出變數設定為類別變數，有 6 個程度(Level)，分別為 0,1,2,3,4,5，其中類別 0(Class 0)代表該路段未發生事故，而類別 1(Class 1)則表示該路段每年發生 1 次事故，類別 2(Class 2)代表該路段每年發生事故在 2 次，類別 3(Class 3)代表該路段每年發生事故在 3~4 次，類別 4(Class 4)代表該路段每年發

生事故在 5~8 次，類別 5(Class 5)代表該路段每年發生事故在 9 次以上(包含 9 次)，其各程度等級如下圖所示：

0	1	2	3	4	5
安全	普通 安全	不安 全	普通 危險	危險	極 危險

圖 6.1 肇事危險程度等級

而決策樹所產生的規則並非全部都是好的，本文將根據[Michael J. A. Berry, 1997]所提出的三種方法，衡量決策樹的有效程度：

- 1.進入該節點之資料數，設定為至少 30 筆。
- 2.若是葉部節點，將觀察資料分類之方式。
- 3.該節點將資料正確分類的比率，設定為至少達 50%。

本研究為評估決策樹的分類效益，採用錯誤率與分類結果矩陣表呈現分類結果，對於不良的決策樹結果，將進行樹的深度、節點純度及節點數之調整，重新訓練樹，直到萃取出有意義的規則。

### 6.3 分類與迴歸樹建立與分析

本研究利用敏感度分析及彈性分析篩選出顯著變數，透過 C 4.5 軟體之 Gini 及 Twoing 來分析路口行向肇事之危險程度，而 CART 分析時必須以學習樣本(Learning Set)建立樹狀模式，再透過測試樣本(Testing Set)來修剪樹，所以必須將事故資料分成訓練組與測試組，由於若採用隨機選取的方式，會使得每次分析的結果不一致，故本研究採以與類神經網路相同之訓練與驗證筆數。即為將事故資料中 779 筆先進行各模式之訓練與學習，260 筆資料對於所建立之模式予以驗證。輸出的目標變數(Target Variable)為肇事危險程度，將此變數設定為類別變數，有 6 個程度(Level)，分別為 0,1,2,3,4,5，其中類別 0(Class



0)代表路口行向為安全，類別 1(Class 1)則表示路口行向為普通安全，類別 2(Class 2)代表路口行向為普通不安全，類別 3(Class 3)則表示路口行向為普通危險，類別 4(Class 4)則表示路口行向為危險，類別 5(Class 5)則表示路口行向為極危險。而預測變數(Predictors)，除路口總交通量不考慮行向組合外，其他依篩選後之 10 個變數，各依行向組合之左轉車輛比、道路寬度、機車比、行向流量、速限或警告標誌、時相複雜程度、時制週期、左/右轉車輛比及快慢車分隔形式共輸入 19 個變數。各變數的符號整理如表 6.2 所示：

表 6.2 決策樹輸入變數代號

代號	名稱/描述/類型
Danger	目標變數/0：0,1：1,2：2,3：3~4,4：5~8,5：9+/類別型
Lane_Partition	快慢車道分隔型式/0：無，1：實體,2：反光，3：標線/類別型
Speed_Limit	速限或警告標誌/0：無，1：有/類別型
Road_Width	道路寬度/連續型
Total_AADT	總交通量/連續型
AADT	行向交通量/連續型
Left_Vehicle_Rate	左轉車輛比/連續型
Right_Vehicle_Rate	右轉車輛比/連續型
Motorcycle_Rate	機車比/連續型
Phase	時相複雜程度/連續型
Cycle	時制週期/連續型

決策樹建立分別以 Gini 與 Twoing 方式，皆過資料比率優先(Priority by Data)之方式建立，下面將分別說明其分析結果。

### 6.3.1 Gini 結果

決策樹會產生 47 個終端節點(Terminal Node)，其結構如圖 6.2 所示(詳圖請參閱附錄 A)，第一個節點(Node 1)的分割條件為道路寬度

<14.25 米(參閱附錄 A)，且第二個節點(Node 2)的分割條件為行向流量 <633.5 pcu 時，左子節點會產生一個終端節點 1(TN1)，表示此 20 個路口中路口為安全之類別比率最高，因此在道路寬度 <14.25 米，且行向流量 <633.5 pcu 時，在決策樹中判斷為安全之路口，右子節點(Node 4)有 488 個路口，再透過行向流量  $\geq 1,828$  pcu 且機車比率 <0.029(參閱附錄 A)，分割子出節點，以此分割至終端節點。

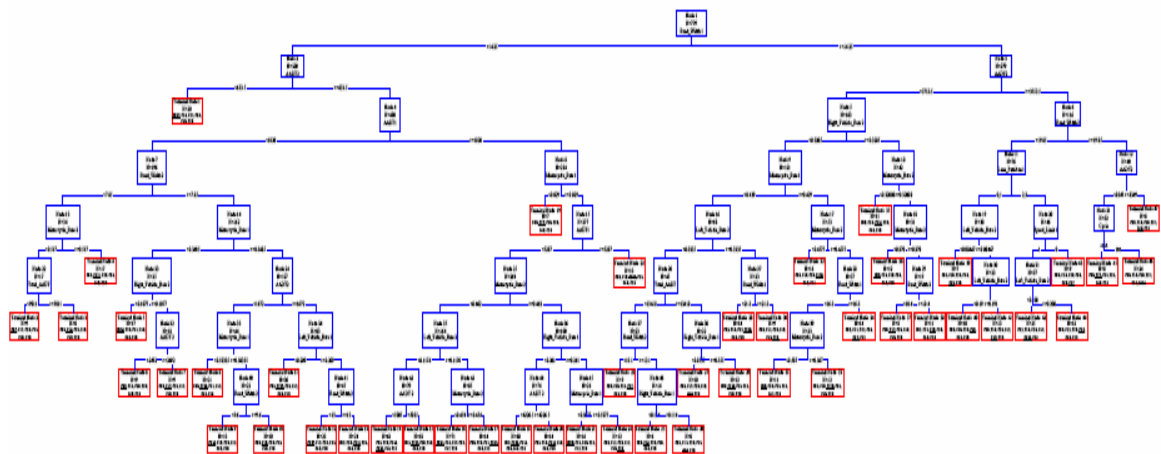


圖 6.2 Gini 決策樹

每個終端節點皆會有最高比率的肇事次數類別，依此最多的類別表示該節點所分類的結果，各個類別與終端節點彙整如表 6.3：該表中 TN42、TN43 及 TN44 為路口危險程度為普通危險至極危險等級。相同的判斷條件為甲方行向於道路寬度  $\geq 14.25$  米、乙方行向流量  $\geq 3753.5$  pcu、乙方行向於道路寬度 <19.85 米與甲方行向於快慢車道間分隔型式以反光標鈕分隔或標線分隔下，路口已經達到普通危險等級以上。另外，當甲方行向路口具有速限或警告標誌時，不至於令路口危險等級達到極危險。再透過乙方行向之左轉車輛比率 <0.208，該路口危險等級為危險；其左轉車輛比率  $\geq 0.208$  時，則路口危險等級為普通危險。

表 6.3 Gini 決策樹類別與終端節點表

危險等級	終端節點編號(Terminal Node)
安全(0)	TN1, TN2, TN5, TN6, TN9, TN12, TN16, TN21, TN31, TN36
普通安全(1)	TN3, TN4, TN7, TN8, TN10, TN11, TN15, TN18, TN19, TN25, TN29, TN37, TN45
普通不安全(2)	TN13, TN20, TN22, TN33, TN35, TN38
普通危險(3)	TN14, TN17, TN23, TN28, TN32, TN40, TN44
危險(4)	TN26, TN27, TN34, TN39, TN41, TN42, TN47
極危險(5)	TN24, TN30, TN43, TN46

6.3.2 Twoing 結果

決策樹會產生 27 個終端節點(Terminal Node)，其結構如圖 6.3 所示(詳圖請參閱附錄 B)，Twoing 分類情形大致與 Gini 差不多，其第一個節點(Node 1)的分割條件亦為道路寬度<14.25 米(參閱附錄 B)。

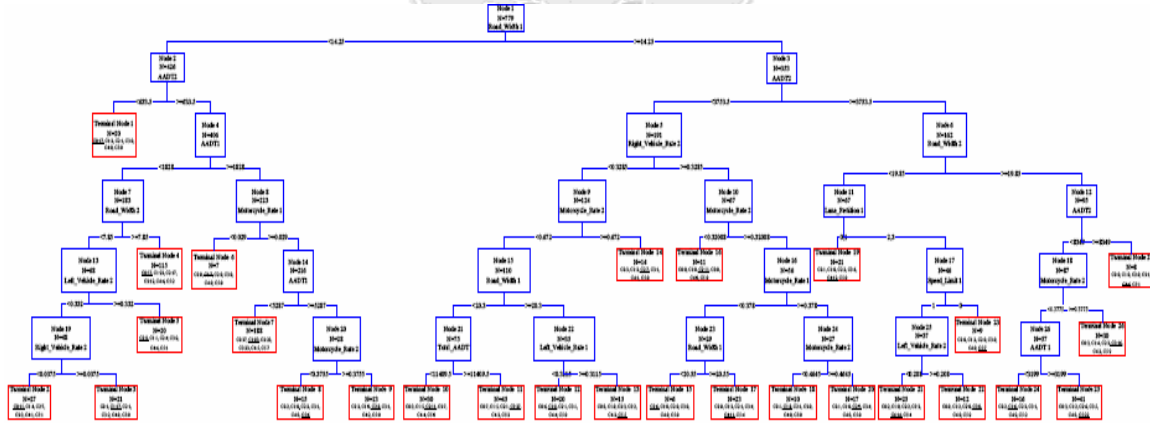


圖 6.3 Twoing 決策樹

各個類別與終端節點彙整如表 6.4，該表得知 Twoing 分類等級為極危險之終端節點不變外，其餘判斷路口危險程度之分類亦明顯減少，較 Gini 分類法差。

表 6.4 Twoing 決策樹類別與終端節點表

危險等級	終端節點編號(Terminal Node)
安全(0)	TN1, TN2, TN4, TN5, TN15
普通安全(1)	TN3, TN6, TN7, TN12, TN18, TN24
普通不安全(2)	TN9, TN10, TN14, TN16, TN20
普通危險(3)	TN11, TN22, TN26
危險(4)	TN17, TN19, TN21, TN27
極危險(5)	TN8, TN13, TN23, TN25

### 6.3.3 預測正確率比較

本研究針對所建構之 Gini 與 Twoing 決策樹進行模式驗證(表 6.5 及 6.6)，成果於危險程度判斷準則總正確率中以 Gini 之總預測較高，為 69.96%(訓練組)、68.08%(驗證組)，明顯高於 Twoing 之 52.37%(訓練組)、48.85%(驗證組)。而從個別正確率可發現危險等級為 5 於兩方法中，訓練組皆具相同之預測結果，顯然對於較危險之路口，其判斷準則一致。該危險程度等級為 0 及 1 的判定，Gini 之正確率近 80% 準確，而 Twoing 卻僅有 60% 上下；然危險等級於 2~4 之判定，於 Gini 均亦有 60% 以上之準確度，而 Twoing 準確率卻不及 50%。從任何角度觀察，Gini 之判定準則優於 Twoing，故本研究採以 Gini 之判定準則作為行向路口危險程度之判定依據。

表 6.5 Gini 決策樹預測正確率表

預測 實際	訓練組								驗證組							
	0	1	2	3	4	5	列總和	正確率 (%)	0	1	2	3	4	5	列總和	正確率 (%)
0	142	14	5	8	7	0	176	80.68	44	5	3	3	1	0	56	78.57
1	10	151	1	10	4	8	184	82.07	4	49	6	0	2	1	62	79.03
2	13	24	69	14	5	0	125	60.00	6	3	22	5	1	0	37	59.46
3	15	17	6	91	9	2	140	65.00	4	7	2	29	3	2	47	61.70
4	13	7	2	8	52	4	86	60.05	1	2	2	5	21	4	35	60.00
5	6	7	2	5	8	40	68	58.82	0	2	1	3	5	12	23	52.17
行總和	199	220	85	136	85	54	779		59	68	36	45	33	19	260	
總正 確率	69.96%								68.08%							

表 6.6 Twoing 決策樹預測正確率表

預測 實際	訓練組								驗證組							
	0	1	2	3	4	5	列總和	正確率 (%)	0	1	2	3	4	5	列總和	正確率 (%)
0	99	47	9	10	3	8	176	56.25	31	13	7	3	0	2	56	55.36
1	30	129	7	11	3	4	184	70.11	8	41	4	5	2	2	62	66.13
2	29	29	46	4	5	12	125	36.80	7	6	12	5	3	4	37	32.43
3	17	37	13	51	14	8	140	36.43	6	9	8	16	5	3	47	34.04
4	9	11	10	5	43	8	86	50.00	1	5	6	4	16	3	35	45.71
5	4	9	0	6	9	40	68	58.82	1	3	0	4	4	11	23	47.83
行總 和	188	262	85	87	77	80	779		54	77	37	37	30	25	260	
總正 確率	52.37%								48.85%							

### 6.3.4 變數重要度

重要性變數之選擇是以節點分支之根部節點為最重要變數 (Primary Variable)，再者從根部往下切割時，主要被選擇出來作為某節點分割條件的主要分割者，假使某個別觀測者之資料遺失而透過代理者代替，使得代理者和主要分割者為某節點分割條件的地位是相等的，此代理者也會成為本模式中之重要性變數，透過第三章 3-21 及 3-22 公式得到將某節點中各候選者能改善此節點純度之改善值 (Improvement Weight)，最後將這些主要分割者與代理者加總得出其重要性變數之分數值。

由表 6.7 發現，二者演算法之變數重要性之排序一致，而 Twoing 變數的重要性比 Gini 變數的重要性還來得少，其二者最主要變數是以道路寬度為主，其分數為 100，再者是行向流量為第二重要變數中 Gini 的分數比 Twoing 高，但分數差距並不會很多，其次為機車比、左/右轉車輛比、路口總交通量、快慢車分隔形式及速限或警告標誌。

表 6.7 Gini 與 Twoing 變數重要度

順序	Gini		Twoing	
	變數	分數	變數	分數
1	道路寬度	100.000	道路寬度	100.000
2	行向流量	79.513	行向流量	74.351
3	機車比	67.467	機車比	64.168
4	左轉車輛比	57.544	左轉車輛比	52.194
5	右轉車輛比	52.647	右轉車輛比	42.427
6	路口總交通量	33.264	路口總交通量	29.466
7	快慢車分隔形式	14.223	快慢車分隔形式	13.331
8	速限或警告標誌	5.377	速限或警告標誌	5.314
9	時制週期	4.631	時制週期	0.000
10	時相複雜程度	0.000	時相複雜程度	0.000

## 6.4 Gini 決策樹實證分析

本研究相較二者決策樹分類法，發現 Gini 分類法是以「節點純度」之角度將父節點分割出兩個子節點，以致於分類錯誤能降低，分割出來的結果可信度也愈高；Twoing 則是以「二分法」，主要為計算父節點與子節點間最大的不純度改善值。但決策樹建立完成後發現，雖二者在判斷法則上差異甚小，而可判斷之類別數多寡及預測準確度皆以 Gini 決策樹較佳。因此，本研究採以 Gini 決策樹為判定路口危險安全程度之準則。為了讓決策者易於瞭解其法則，下列將針對 Gini 決策樹節點與判斷法則測試做細部說明。

### 6.4.1 Gini 決策樹節點分析

決策樹產生 47 個終端節點(Terminal Node)，其結構如圖 6.2 所示(詳圖請參閱附錄 A)，各節點皆以預測變數為此點之判斷指標，如 Road\_Width 1(Node 1)、AADT 2(Node 2、3)及 Right\_Vehicle\_Rate 2(Node 5)。本研究之研究範圍為路口之交通事故，而每件事故是依照車進入路口之上游路段之行向記錄，如一十字路口東西向為紅燈，西向車輛闖紅燈撞上南向車輛，而涉及東向停等車輛，則記為西向與南向之肇事。其中行向「西向東」係指節點之「預測變數 1」，亦稱為「甲方」；行向「南向」則為「預測變數 2」亦為「乙方」。

而節點間具有父子點之關係。父點與子點所分割之觀測值，為透過本研究針對預測變數給予之參數做假設，利用 C 4.5 軟體切割所得。下列將介紹觀測值於節點所代表之涵義，以節點—父點之預測變數(觀測值)表示。

#### 1. Node 1—Road\_Width 1(14.25 米)

道路寬度為決策樹首要考量之因素，係指判定路口是否存在危險之重要因素。目前一般車道寬度為 3.5 米，若依決策樹所得之觀測值 14.25 米劃分，係為四個快車道，而因台灣地區機車數量過多，故有機車道或者混合車道。依目前規劃之車道型式，道路寬度於 14.25 米內之車道型式有單行道、雙向道、雙向二車道及雙向三車道；而超過 14.25 米之車道型式則有雙向四車道、雙向六車道及雙向八車道等。

## 2. Node 2 – AADT 2(633.5 pcu)

為 Node 1 之子點，係指判定路口是否存在危險，在考量道路寬度小於 14.25 米後，仍須考慮年平均日交通量。而此時觀測值為 633.5 pcu，即為小於 633.5 pcu 時，此路口為安全路口(TN 1)；超過 633.5 pcu 時，則必須再透過其他子點之觀測值來衡量，才能判定路口是否存在危險。

## 3. Node 3 – AADT 2(3,753.5 pcu)

亦為 Node 1 之子點，但其為考量道路寬度大於 14.25 米後，再以年平均日交通量於 3,753.5pcu 為切割，並仍須考慮年平均日交通量。其後之判斷準則皆同於上述。

經簡述分析後發現，欲瞭解路口是否存在危險，非僅為某單一因素之考量，應透過其他因素之限制而形成法則，本研究透過 Gini 分割出 47 個終端節點，即表示有 47 條判斷路口安全之法則，而各影響路口安全之法則如下：

表 6.8 Gini 決策樹判斷路口安全危險程度法則

路口危險程度	法則	判斷路口安全之法則
安全	1	Road_Width 1 <14.25 米，AADT 2 <633.5 pcu
	2	Road_Width 1 <14.25 米，AADT 2 $\geq$ 633.5 pcu，AADT 1 <1,828 pcu，Road_Width 2 <7.85 米，Motorcycle_Rate 2 <0.517，Total_AADT <9,811 pcu
	3	Road_Width 1 <14.25 米，AADT 2 $\geq$ 633.5 pcu，AADT 1 <1,828 pcu，Road_Width $\geq$ 7.85 米，Motorcycle_Rate 1 <0.3485，Right_Vehicle_Rate 2 <0.1077
	4	Road_Width 1 <14.25 米，AADT 2 $\geq$ 633.5 pcu，AADT 1 <1,828 pcu，Road_Width 2 $\geq$ 7.85 米，Motorcycle_Rate 1 <0.3485，Right_Vehicle_Rate 2 $\geq$ 0.1077，AADT 2 <2,492 pcu



表 6.8 Gini 決策樹判斷路口安全危險程度法則(續)

路口危險程度	法則	判斷路口安全之法則
安全	5	Road_Width 1 <14.25 米 , AADT 2 $\geq$ 633.5 pcu , AADT 1 <1,828 pcu , Road_Width 2 $\geq$ 7.85 米 , Motorcycle_Rate 1 $\geq$ 0.3485 , AADT 2 <1,672 pcu , Motorcycle_Rate 1 $\geq$ 0.50335 , Road_Width 2 <9.6 米
	6	Road_Width 1 <14.25 米 , AADT 2 $\geq$ 633.5 pcu , AADT 1 <1,828 pcu , Road_Width 2 $\geq$ 7.85 米 , Motorcycle_Rate 1 $\geq$ 0.3485 , AADT 2 $\geq$ 1,672 pcu , Left_Vehicle_Rate 1 $\geq$ 0.209 , Road_Width 2 <13 米
	7	Road_Width 1 <14.25 米 , AADT 2 $\geq$ 633.5 pcu , AADT 1 $\geq$ 1,828 pcu , Motorcycle_Rate 1 $\geq$ 0.029 , AADT 1 <5,287 pcu , Motorcycle_Rate 2 <0.462 , Left_Vehicle_Rate 2 $\geq$ 0.1176 , Motorcycle_Rate 2 <0.454
	8	Road_Width 1 <14.25 米 , AADT 2 $\geq$ 633.5 pcu , AADT 1 $\geq$ 1,828 pcu , Motorcycle_Rate 1 $\geq$ 0.029 , AADT 1 <5,287 pcu , Motorcycle_Rate 2 $\geq$ 0.462 , Right_Vehicle_Rate 1 $\geq$ 0.261 , Motorcycle_Rate 1 <0.5075
	9	Road_Width 1 $\geq$ 14.25 米 , AADT 2 <3,753.5 pcu , Right_Vehicle_Rate 2 <0.3285 , Motorcycle_Rate 1 $\geq$ 0.429 , Motorcycle_Rate 2 $\geq$ 0.4575 , Road_Width 1 <16.3 米 , Motorcycle_Rate 2 <0.507
	10	Road_Width 1 $\geq$ 14.25 米 , AADT 2 <3,753.5 pcu , Right_Vehicle_Rate 2 $\geq$ 0.3285 , Motorcycle_Rate 2 $\geq$ 0.32008 , Motorcycle_Rate 2 <0.378

表 6.8 Gini 決策樹判斷路口安全危險程度法則(續)

路口危險程度	法則	判斷路口安全之法則
普通安全	11	Road_Width 1 <14.25 米, AADT 2 $\geq$ 633.5 pcu, AADT 1 <1,828 pcu, Road_Width 2 <7.85 米, Motorcycle_Rate 2 <0.517, Total_AADT $\geq$ 9,811 pcu
	12	Road_Width 1 <14.25 米, AADT 2 $\geq$ 633.5 pcu, AADT 1 <1,828 pcu, Road_Width 2 <7.85 米, Motorcycle_Rate 2 $\geq$ 0.517
	13	Road_Width 1 <14.25 米, AADT 2 $\geq$ 633.5 pcu, AADT 1 <1,828 pcu, Road_Width 2 $\geq$ 7.85 米, Motorcycle_Rate 1 <0.3485, Right_Vehicle_Rate 2 >0.1077, AADT 2 $\geq$ 2,492 pcu
	14	Road_Width 1 <14.25 米, AADT 2 $\geq$ 633.5 pcu, AADT 1 <1,828 pcu, Road_Width 2 $\geq$ 7.85 米, Motorcycle_Rate 1 $\geq$ 0.3485, AADT 2 <1,672 pcu, Motorcycle_Rate 1 <0.50335
	15	Road_Width 1 <14.25 米, AADT 2 $\geq$ 633.5 pcu, AADT 1 <1,828 pcu, Road_Width 2 $\geq$ 7.85 米, Motorcycle_Rate 1 $\geq$ 0.3485, AADT 2 <1,672 pcu, Motorcycle_Rate 1 $\geq$ 0.50335, Road_Width 2 $\geq$ 9.6 米
	16	Road_Width 1 <14.25 米, AADT 2 $\geq$ 633.5 pcu, AADT 1 <1,828 pcu, Road_Width 2 $\geq$ 7.85 米, Motorcycle_Rate 1 $\geq$ 0.3485, AADT 2 $\geq$ 1,672 pcu, Left_Vehicle_Rate 1 <0.209
	17	Road_Width 1 <14.25 米, AADT 2 $\geq$ 633.5 pcu, AADT 1 $\geq$ 1,828 pcu, Motorcycle_Rate 1 <0.029
	18	Road_Width 1 <14.25 米, AADT 2 $\geq$ 633.5 pcu, AADT 1 $\geq$ 1,828 pcu, Motorcycle_Rate 1 $\geq$ 0.029, AADT 1 <5,287 pcu, Motorcycle_Rate 2 <0.462, Left_Vehicle_Rate 2 <0.1176, AADT 2 $\geq$ 5,005 pcu

表 6.8 Gini 決策樹判斷路口安全危險程度法則(續)

路口危險程度	法則	判斷路口安全之法則
普通安全	19	Road_Width 1 <14.25 米, AADT 2 $\geq$ 633.5 pcu, AADT 1 $\geq$ 1,828 pcu, Motorcycle_Rate 1 $\geq$ 0.029, AADT 1 <5,287 pcu, Motorcycle_Rate 2 $\geq$ 0.462, Right_Vehicle_Rate 1 <0.261, AADT 2 <4,226.5 pcu
	20	Road_Width 1 $\geq$ 14.25 米, AADT 2 <3,753.5 pcu, Right_Vehicle_Rate 2 <0.3285, Motorcycle_Rate 1 <0.429, Left_Vehicle_Rate 2 <0.3335, Total_AADT <13,418 pcu, Road_Width 2 $\geq$ 13.1 米, Right_Vehicle_Rate 1 <0.114
	21	Road_Width 1 $\geq$ 14.25 米, AADT 2 <3,753.5 pcu, Right_Vehicle_Rate 2 <0.3285, Motorcycle_Rate 1 <0.429, Left_Vehicle_Rate 2 <0.3335, Total_AADT $\geq$ 13,418 pcu, Right_Vehicle_Rate 1 $\geq$ 0.555
	22	Road_Width 1 $\geq$ 14.25 米, AADT 2 <3,753.5 pcu, Right_Vehicle_Rate 2 $\geq$ 0.3285, Motorcycle_Rate 2 $\geq$ 0.32008, Motorcycle_Rate 2 $\geq$ 0.378, Road_Width 2 <10.4 米
	23	Road_Width 1 $\geq$ 14.25 米, AADT 2 $\geq$ 3,753.5 pcu, Road_Width 2 $\geq$ 19.85 米, AADT 2 <8,349 pcu, Cycle <155 秒
普通不安全	24	Road_Width 1 <14.25 米, AADT 2 $\geq$ 633.5 pcu, AADT 1 <1,828 pcu, Road_Width 2 $\geq$ 7.85 米, Motorcycle_Rate 1 $\geq$ 0.3485, AADT 2 $\geq$ 1,672 pcu, Left_Vehicle_Rate 1 $\geq$ 0.209, Road_Width 2 $\geq$ 13.1 米
	25	Road_Width 1 <14.25 米, AADT 2 $\geq$ 633.5 pcu, AADT 1 $\geq$ 1,828 pcu, Motorcycle_Rate 1 $\geq$ 0.029, AADT 1 $\geq$ 5,287 pcu

表 6.8 Gini 決策樹判斷路口安全危險程度法則(續)

路口危險程度	法則	判斷路口安全之法則
普通 不安 全	26	Road_Width 1 <14.25 米, AADT 2 $\geq$ 633.5 pcu , AADT 1 $\geq$ 1,828 pcu , Motorcycle_Rate 1 $\geq$ 0.029 , AADT 1 <5,287 pcu , Motorcycle_Rate 2 $\geq$ 0.462 , Right_Vehicle_Rate 1 <0.261 , AADT 2 $\geq$ 4,226.5 pcu
	27	Road_Width 1 $\geq$ 14.25 米 , AADT 2 <3,753.5 pcu , Right_Vehicle_Rate 2 <0.3285 , Motorcycle_Rate 1 $\geq$ 0.429 , Motorcycle_Rate 2 $\geq$ 0.4575 , Road_Width 1 <16.3 米 , Motorcycle_Rate 2 $\geq$ 0.507
	28	Road_Width 1 $\geq$ 14.25 米 , AADT 2 <3,753.5 pcu , Right_Vehicle_Rate 2 $\geq$ 0.3285 , Motorcycle_Rate 2 <0.32008
	29	Road_Width 1 $\geq$ 14.25 米 , AADT 2 <3,753.5 pcu , Right_Vehicle_Rate 2 $\geq$ 0.3285 , Motorcycle_Rate 2 $\geq$ 0.32008 , Motorcycle_Rate 2 $\geq$ 0.378 , Road_Width 2 $\geq$ 10.4 米
普通 危 險	30	Road_Width 1 <14.25 米 , AADT 2 $\geq$ 633.5 pcu , AADT 1 $\geq$ 1,828 pcu , Motorcycle_Rate 1 $\geq$ 0.029 , AADT 1 <5,287 pcu , Motorcycle_Rate 2 <0.462 , Left_Vehicle_Rate 2 <0.1176 , AADT 2 <5,005 pcu
	31	Road_Width 1 <14.25 米 , AADT 2 $\geq$ 633.5 pcu , AADT 1 $\geq$ 1,828 pcu , Motorcycle_Rate 1 $\geq$ 0.029 , AADT 1 <5,287 pcu , Motorcycle_Rate 2 <0.462 , Left_Vehicle_Rate 2 $\geq$ 0.1176 , Motorcycle_Rate 2 $\geq$ 0.454
	32	Road_Width 1 $\geq$ 14.25 米 , AADT 2 <3,753.5 pcu , Right_Vehicle_Rate 2 <0.3285 , Motorcycle_Rate 1 <0.429 , Left_Vehicle_Rate 2 <0.3335 , Total_AADT <13,418 pcu , Road_Width 2 <13.1 米

表 6.8 Gini 決策樹判斷路口安全危險程度法則(續)

路口危險程度	法則	判斷路口安全之法則
普通危險	33	Road_Width 1 $\geq 14.25$ 米 , AADT 2 $< 3,753.5$ pcu , Right_Vehicle_Rate 2 $< 0.3285$ , Motorcycle_Rate 1 $< 0.429$ , Left_Vehicle_Rate 2 $\geq 0.3335$ , Road_Width 1 $< 21.3$ 米
	34	Road_Width 1 $\geq 14.25$ 米 , AADT 2 $< 3,753.5$ pcu , Right_Vehicle_Rate 2 $< 0.3285$ , Motorcycle_Rate 1 $\geq 0.429$ , Motorcycle_Rate 2 $< 0.4575$
	35	Road_Width 1 $\geq 14.25$ 米 , AADT 2 $\geq 3,753.5$ pcu , Road_Width 2 $< 19.85$ 米 , Lane_Partition 1 =0,1 , Left_Vehicle_Rate 2 $\geq 0.08447$ , Left_Vehicle_Rate 2 $< 0.191$
	36	Road_Width 1 $\geq 14.25$ 米 , AADT 2 $\geq 3,753.5$ pcu , Road_Width 2 $< 19.85$ 米 , Lane_Partition 1 =2,3 , Speed_Limit 1 =1 , Left_Vehicle_Rate 2 $\geq 0.208$
危險	37	Road_Width 1 $\geq 14.25$ 米 , AADT 2 $< 3,753.5$ pcu , Right_Vehicle_Rate 2 $< 0.3285$ , Motorcycle_Rate 1 $< 0.429$ , Left_Vehicle_Rate 2 $< 0.3335$ , Total_AADT $< 13,418$ pcu , Road_Width 2 $\geq 13.1$ 米 , Right_Vehicle_Rate 1 $\geq 0.114$
	38	Road_Width 1 $\geq 14.25$ 米 , AADT 2 $< 3,753.5$ pcu , Right_Vehicle_Rate 2 $< 0.3285$ , Motorcycle_Rate 1 $< 0.429$ , Left_Vehicle_Rate 2 $< 0.3335$ , Total_AADT $\geq 13,418$ pcu , Right_Vehicle_Rate 1 $< 0.555$
	39	Road_Width 1 $\geq 14.25$ 米 , AADT 2 $< 3,753.5$ pcu , Right_Vehicle_Rate 2 $< 0.3285$ , Motorcycle_Rate 1 $\geq 0.429$ , Motorcycle_Rate 2 $\geq 0.4575$ , Road_Width 1 $\geq 16.3$ 米

表 6.8 Gini 決策樹判斷路口安全危險程度法則(續)

路口危險程度	法則	判斷路口安全之法則
危險	40	Road_Width 1 $\geq$ 14.25 米 , AADT 2 $\geq$ 3,753.5 pcu , Road_Width 2 < 19.85 米 , Lane_Partition 1 = 0, 1 , Left_Vehicle_Rate 2 < 0.08447
	41	Road_Width 1 $\geq$ 14.25 米 , AADT 2 $\geq$ 3,753.5 pcu , Road_Width 2 < 19.85 米 , Lane_Partition 1 = 0, 1 , Left_Vehicle_Rate 2 $\geq$ 0.08447 , Left_Vehicle_Rate 2 $\geq$ 0.191
	42	Road_Width 1 $\geq$ 14.25 米 , AADT 2 $\geq$ 3,753.5 pcu , Road_Width 2 < 19.85 米 , Lane_Partition 1 = 2, 3 , Speed_Limit 1 = 1 , Left_Vehicle_Rate 2 < 0.208
	43	Road_Width 1 $\geq$ 14.25 米 , AADT 2 $\geq$ 3,753.5 pcu , Road_Width 2 $\geq$ 19.85 米 , AADT 2 $\geq$ 8,349 pcu
極危險	44	Road_Width 1 < 14.25 米 , AADT 2 $\geq$ 633.5 輛 , AADT 1 $\geq$ 1,828 pcu , Motorcycle_Rate 1 $\geq$ 0.029 , AADT 1 < 5,287 pcu , Motorcycle_Rate 2 $\geq$ 0.462 , Right_Vehicle_Rate 1 $\geq$ 0.261 , Motorcycle_Rate 1 $\geq$ 0.5075
	45	Road_Width 1 $\geq$ 14.25 米 , AADT 2 < 3,753.5 pcu , Right_Vehicle_Rate 2 < 0.3285 , Motorcycle_Rate 1 < 0.429 , Left_Vehicle_Rate 2 $\geq$ 0.3335 , Road_Width 1 $\geq$ 21.3 米
	46	Road_Width 1 $\geq$ 14.25 米 , AADT 2 $\geq$ 3,753.5 pcu , Road_Width 2 < 19.85 米 , Lane_Partition 1 = 2, 3 , Speed_Limit 1 = 0
	47	Road_Width 1 $\geq$ 14.25 米 , AADT 2 $\geq$ 3,753.5 pcu , Road_Width 2 $\geq$ 19.85 米 , AADT 2 < 8,349 pcu , Cycle = 180 秒

由表 6.8 發現，判定路口為安全之法則有 10 條；普通安全有 13 條；普通不安全有 6 條；普通危險有 7 條；危險有 7 條，最後極危險則有 4 條。法則之劃分之差異，主要受限於本研究資料蒐集時，行向組合之交通事故資料以發生次數為一次及零次之行向組合總數最多，其後為隨之遞減之緣故。但各法則在判斷過程皆以道路寬度之觀測值 14.25 米為優先準則。發現道路寬度小於 14.25 米中，判定路口為安全之法則有 8 條；普通安全有 9 條；普通不安全有 3 條；普通危險有 2 條；危險有 0 條，最後極危險則有 1 條；相較其他法則發現，路口程度愈為危險的法則皆多數為道路寬度大於 14.25 米，故顯然道路寬度愈寬，路口危險程度會愈嚴重。為了探討道路寬度之觀測值 14.25 米是否為參考依據，本研究以實例事故案例作測試，希望能有好的結果。

#### 6.4.2 實例測試

依本研究假設各臨近路段進入路口，其面對之道路環境均有所不同，而對肇事有所影響下，並考量上述之 47 個法則是否適應肇事路口主要肇因之判定，將以五權西路與環中路路口之不同行向組合所發生事故次數，觀察其法則之適用性。其基本調查資料與路口示意圖如表 6.9 與圖 6.4 所示：

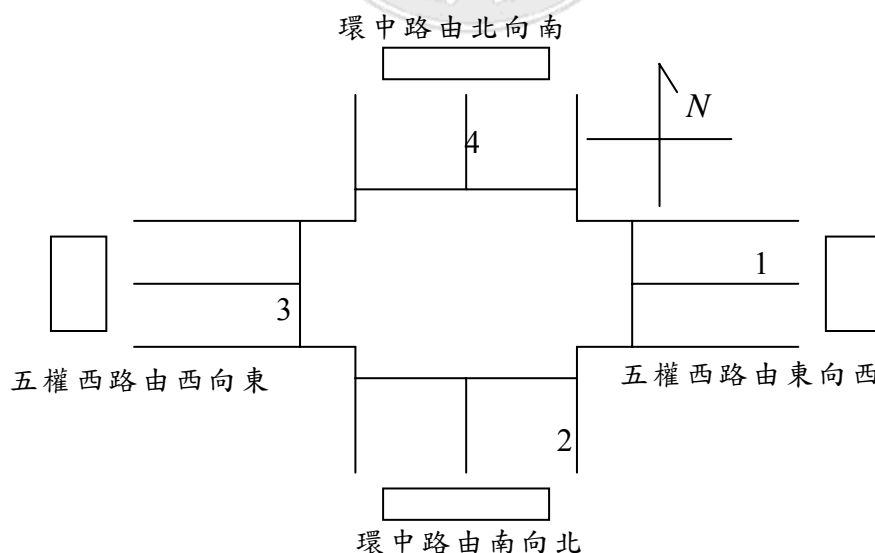


圖 6.4 五權西路與環中路交叉路口示意圖

表 6.9 五權西路與環中路路口之基本調查資料

時間	民國 92、93 年			
行向代碼	1	2	3	4
道路名稱	五權西路	環中路	五權西路	環中路
車輛行向	東	南	西	北
快車道數	3	4	3	4
中央分隔	實體分隔	實體分隔	實體分隔	實體分隔
快慢分隔	標線	實體分隔	標線	實體分隔
左轉專用道	1	1	1	1
道路寬度(m)	13.3	26.5	12.4	26.2
道路被佔用程度	未遭佔用	未遭佔用	未遭佔用	未遭佔用
時相複雜程度	4	4	4	4
號誌個數(個)	5	5	5	5
左轉用號誌	有	有	有	有
號誌周期	180	180	180	180
綠燈秒數(秒)	78	32	78	32
行向流量(pcu)	3,961	3,396	5,392	4,665
機車比	0.551	0.354	0.348	0.468
左轉比	0.167	0.378	0.202	0.303
右轉比	0.341	0.256	0.275	0.355
限速或警告標誌	無	有	無	有
十公尺內是否有障礙物	否	是	否	是

各行向組合事故次數如表 6.10。針對此路口，本研究採以行向一與行向二、行向一與行向四、行向一與行向三、行向二與行向四、行向三與行向三及行向四與行向四進行分析。

表 6.10 五權西路與環中路路口各行向組合之交通事故實際次數

行向組合	事故次數	行向組合	事故次數
	92、93 年		92、93 年
1&1	2	2&3	7
1&2	0	2&4	15
1&3	1	3&3	2
1&4	12	3&4	5
2&2	21	4&4	17



由表 6.11 與 6.12 知，肇事型態皆為橫向之碰撞事故，分別為行向一與行向二及行向一與行向四，發生之事故次數各為 0 次與 12 次。為瞭解相同路口下，肇事次數卻有如此差異之情況，將各行向之基本調查資料輸入，透過 Gini 決策樹之 47 條法則發現，行向一與行向二之行向組合結果為安全路口(法則 7)，而行向一與行向四為極危險路口(法則 44)。二者在判斷之過程中，發現幾何設計差異不大，以交通動態為主要原因。發現行向四之流量較行向二多，故此路段之道路容量較為擁擠，且機車比率亦較多之情況下，機車駕駛者因心情較浮躁，以蛇行行駛強行通過，此時汽車駕駛者稍不注意，即釀成事故發生。

實際上，此路口為通往工業區必經孔道，而由彰化等地至台中，或中山高速公路北上中途下車前往台中市區之車輛往往經南屯交流道北上匝道，再轉五權西路經環中路轉往其目的地，由於流量往往超過原先規劃之道路容量，以致經常發生阻塞。而在上下班的尖峰時段，紅燈違規右轉、隨意變換車道及位號誌行駛的車輛比比皆是，是故本研究之法則與實際情況符合。

表 6.11 輸入變數與決策樹輸出結果資料對照表(行向一與行向二)

因素	輸入變數		決策樹判斷結果	
	東向西	南向北	東向西	南向北
道路寬度(m)	13.3	26.5	<14.25	
行向流量(pcu)	3,961	3,396	$\geq 1,828$ <5,287	$\geq 633.5$
機車比	0.551	0.354	$\geq 0.029$	<0.462 <0.454
左轉車輛比	0.167	0.378		$\geq 0.1176$
右轉車輛比	0.341	0.256		
路口總交通量	16,714	16,714		
快慢車分隔形式	標線	實體分隔		
速限或警告標誌	無	有		
時制周期	180			
時相複雜程度	4	4		
危險程度	0 次		安全路口	

表 6.12 輸入變數與決策樹輸出結果資料對照表(行向一與行向四)

因素	輸入變數		決策樹判斷結果	
	東向西	北向南	東向西	北向南
道路寬度(m)	13.3	26.2	<14.25	
行向流量(pcu)	3,961	4,665	$\geq 1,828$ <5,287	$\geq 633.5$
機車比	0.551	0.468	$\geq 0.029$ $\geq 0.5075$	$\geq 0.462$
左轉車輛比	0.167	0.303		
右轉車輛比	0.341	0.355	$\geq 0.261$	
路口總交通量	16,714	16,714		
快慢車分隔形式	標線	實體分隔		
速限或警告標誌	無	有		
時制周期	180			
時相複雜程度	4	4		
危險程度	12 次		極危險路口	

由表 6.13 與 6.14 知，肇事型態皆為對向之碰撞事故，分別為行向一與行向三及行向二與行向四，發生之事故次數各為 1 次與 15 次。同上述動作，經由判斷結果發現，行向一與行向三之行向組合結果為普通安全路口(法則 18)，而行向二與行向四為極危險路口(法則 47)。二者在判斷之過程中，發現組合為行向一與行向三之道路寬度皆小於 14.25 米，而受制大量流量及機車比率影響下，影響駕駛者行駛之速度降低，故較少發生事故；而組合為行向二與行向四之道路寬度皆大於 14.25 米，交通動態卻與前述之行向組合差異甚小，反而駕駛者之行駛速度過快，而造成嚴重事故。

實際上，由於外地前往工業區之車流大量湧入，致使五權西路路段過於飽和，而當地上班族欲行駛五權西路前往工業區，必定寸步難行。使得此路段之行駛速度低於 50 km/hr，故此路段較為安全；環中路亦為進入工業區之路段，故流量亦不亞於五權西路，但由於此路段之道路寬度較寬，故往往行經此路段之速度皆為 55 km/hr 以上，甚至超速，故此路段較為嚴重。

表 6.13 輸入變數與決策樹輸出結果資料對照表(行向一與行向三)

因素	輸入變數		決策樹判斷結果	
	東向西	南向北	東向西	南向北
道路寬度(m)	13.3	12.4	<14.25	
行向流量(pcu)	3,961	5,392	$\geq 1,828$ <5,287	$\geq 633.5$ $\geq 5,005$
機車比	0.551	0.348	$\geq 0.029$	<0.462
左轉車輛比	0.167	0.202		<0.1176
右轉車輛比	0.341	0.275		
路口總交通量	16,714	16,714		
快慢車分隔形式	標線	標線		
速限或警告標誌	無	無		
時制周期	180			
時相複雜程度	4	4		
危險程度	1 次		普通安全路口	

表 6.14 輸入變數與決策樹輸出結果資料對照表(行向二與行向四)

因素	輸入變數		決策樹判斷結果	
	南向北	北向南	南向北	北向南
道路寬度(m)	26.5	26.2	$\geq 14.25$	$\geq 19.85$
行向流量(pcu)	3,396	4,665		$\geq 3,753.5$ <8,349
機車比	0.354	0.468		
左轉車輛比	0.378	0.303		
右轉車輛比	0.256	0.355		
路口總交通量	16,714	16,714		
快慢車分隔形式	實體分隔	實體分隔		
速限或警告標誌	有	有		
時制周期	180		180	
時相複雜程度	4	4		
危險程度	15 次		極危險路口	

由表 6.15 與 6.16 知，為行向三與行向三及行向二與行向二之同向肇事型態，事故次數各為 2 次與 21 次。結果發現行向三與行向三之行向組合結果為普通不安全路口(法則 25)，而行向二與行向二為極危險路口(法則 45)。其結果與上述情況大致一樣，主要以道路寬度 14.25 米為分界，受流量及機車比率影響，致使影響駕駛者行車速度。

表 6.15 輸入變數與決策樹輸出結果資料對照表(行向三與行向三)

因素	輸入變數		決策樹判斷結果	
	南向北	南向北	南向北	南向北
道路寬度(m)	12.4	12.4	<14.25	
行向流量(pcu)	5,392	5,392	$\geq 1,828$ $\geq 5,287$	$\geq 633.5$
機車比	0.348	0.348	$\geq 0.029$	
左轉車輛比	0.202	0.202		
右轉車輛比	0.275	0.275		
路口總交通量	16,714	16,714		
快慢車分隔形式	標線	標線		
速限或警告標誌	無	無		
時制周期	180			
時相複雜程度	4	4		
危險程度	2 次		普通不安全路口	

表 6.16 輸入變數與決策樹輸出結果資料對照表(行向二與行向二)

因素	輸入變數		決策樹判斷結果	
	南向北	南向北	南向北	南向北
道路寬度(m)	26.5	26.5	$\geq 14.25$ $\geq 21.3$	
行向流量(pcu)	3,396	3,396		<3753.5
機車比	0.354	0.354	<0.429	
左轉車輛比	0.378	0.378		$\geq 0.3335$
右轉車輛比	0.256	0.256		<0.3285
路口總交通量	16,714	16,714		
快慢車分隔形式	實體分隔	實體分隔		
速限或警告標誌	有	有		
時制周期	180			
時相複雜程度	4	4		
危險程度	21 次		極危險路口	

## 6.5 小結

分類與迴歸樹模式中是以 Gini 決策樹之結果較佳，藉由各行向基本資料之研判，可判斷出碰撞之行向路口存在之危險程度。以道路寬度為 14.25 米作分界。其中於道路寬度 14.25 米內之碰撞事故，安全路口佔 25.2%(126/500)、普通安全路口佔 24.6%(123/500)、普通不安全路口佔 7.8%(39/500)、普通危險路口佔 9.4%(47/500)及極危險路口佔 1.2%(6/500)；而道路寬度 14.25 米外之碰撞事故，安全路口僅佔 5.73%(16/279)、普通安全路口佔 10.04%(28/279)、普通不安全路口佔 10.75%(30/279)、普通危險路口佔 15.77%(44/279)、危險路口佔 18.64%(52/279)及極危險路口佔 12.19%(34/279)。

經實例查證後發現，道路寬度僅為首先判定路口危險程度之參考點。事實上，事故之發生非僅為某單一因素之考量，而是受幾何、號誌及交通狀況因素間之相互影響。一般對於改善交通之手段，大多數針對人為因素、交通動態狀況或者是號誌問題為參考依據，往往忽略幾何設計是否有缺陷。本研究透過分類與迴歸樹結果發現，幾何設計之影響較為交通動態大，其中以道路寬度為首要參考點，因新闢一條新道路時，規劃單位皆會先考量此道路之定位為何，而去規劃道路容量、服務水準及速限，然而一條道路的開闢往往會帶動交通量的隱性需求或都市的發展，導致原本的道路設計容量，漸漸無法滿足實際的流量，因而造成事故頻繁或是交通阻塞嚴重，故時常作號誌調整等動作來改善。目的皆為調解流量的問題，其原因即可能為當初規劃時之道路容量無法滿足現有容量、超過現有容量或是遠少於現有容量，導致駕駛者違規行駛，易造成事故發生，是故道路寬度與行向流量為幾何設計重要之環節。

## 第七章 結論與建議

過去交通肇事分析大多使用統計方法，本研究以類神經網路應用於市區路口事故分析上，並與統計分析方法做比較，透過資料蒐集與整理，模式建立與分析比較後，找出路口事故之相關因素，同時將路口危險程度之判定利用資料探勘給予一套完整之作業規範結果。本章將說明本研究獲得之相關結論及分析結果於台中市市區路口事故的預防，並對未來相關研究提出建議。

### 7.1 結論

1. 在肇事資料中，每年事故發生次數在 12 次以上所佔比率最少，而 1 次與未發生事故之路口行向比率最多，其中以台中港路與文心路發生之 194 次數最多。
2. 本研究以相關係數、誤判率進行之類神經網路與負二項迴歸模式預測能力之比較分析。以相關係數而言，所建構之類神經網路模式為 0.994，負二項迴歸模式為 0.843；誤判率而言，類神經網路模式之誤判率為 13.53%，負二項迴歸模式之誤判率為 57.42%；其模式預測產生之最大誤差而言，類神經網路之最大誤差為 1.37，負二項迴歸模式之最大誤差為 -25.43。由上述得知，不論相關係數與誤差率而言，以類神經網路之預測能力較佳。
3. 由決策樹變數重要程度結果發現，幾何設計之變數影響最大，其次才為交通動態變數。重要排序依次為道路寬度、行向流量、機車比、左/右轉車輛比、路口總交通量、快慢車分隔形式、速限或警告標誌與時制週期等。
4. 各路口行向之判定結果，危險程度為安全路口與普通安全路口在 Gini 決策樹之正確率近達八成；危險程度為普通不安全路口至危險路口，亦有六成以上之準確度；而判定為極危險路口卻僅約五成。其主要受限於本研究資料蒐集時，事故次數以一次及零次佔大多數，比率約佔整體之 23.58%及 22.33%，其後為隨之遞減，而 9 次以上僅約 9.05%，故亦有其代表性。

5. 透過 Gini 決策樹與 Twoing 決策樹分析，以 Gini 決策樹之結果較佳。而以道路寬度 14.25 米作準則。道路寬度 14.25 米內之事故，安全路口佔 25.2%、普通安全路口佔 24.6%、普通不安全路口佔 7.8%、普通危險路口佔 9.4%及極危險路口佔 1.2%；而道路寬度 14.25 米外之事故，安全路口僅佔 5.73%、普通安全路口佔 10.04%、普通不安全路口佔 10.75%、普通危險路口佔 15.77%、危險路口佔 18.64%及極危險路口佔 12.19%。而事故發生所造成之影響程度為何，則透過因素彼此間錯綜複雜之影響。
6. 本研究經由台中市市區之 102 個路口所擬出之路口危險程度之判定模式，此模式是否適應台灣各地，仍有待確認。分類與迴歸樹具有擴建資料之功能，即使有不適之地區，仍可進行修正，加以強化分類之可行性。但國外由於機車數相當稀少，則此模式即不適應。

## 7.2 建議

1. 資料探勘之技術與方法相當多種，本研究採以分類與迴歸樹，未來將繼續此一研究模式，期瞭解路口是否具有危險性，尋找更多之判斷危險程度分類規則，以求決策者規劃有參考依據，以降低路口危險程度之效能。並且希望能嘗試以其他探勘方法，如關聯、預測等，或其他方法，如貝氏分類法，來挖掘其他不同構面上之應用規則。
2. 因本研究所蒐集之資料僅針對道路環境因子與肇事次數間之影響程度，故以肇事次數法作為判斷危險之準則。未來可考慮曝光量變動情形及傷亡嚴重情形，嘗試不同之危險性評估指標，如肇事機率法及事故嚴重性法等，以求辨識路口危險程度更為客觀。
3. 本研究為探討道路環境及交通工程設施因素對於路口事故之嚴重情形，擷取較為重要之 10 種因素進行路口危險程度之判定，事實上，影響交通事故發生之因素尚具人為疏失及天候因素。未來可納入其他因素，再進行決策樹預測，建構出一套路口潛在危險判定系統，使流程簡單化與快速化以加強系統之參考依據。

### 7.3 研究貢獻

本研究主要貢獻在探究道路環境及交通工程設施因素對於路口事故之嚴重情形，針對影響較為重要之因素，本研究選擇適當之資料探勘演算法應用在實際資料中，分析因素間之相互關係，並建立一套路口行向肇事之危險程度規則，獲致初步成果。透過本研究成果之展現，一方面希望幫助決策者於規劃上有參考之應用，另一方面希望能留給未來探勘者一個實作參考，減少摸索嘗試的時間。





## 參考文獻

### 中文部份

1. 交通部網頁， <http://www.motc.gov.tw/>
2. 尤春惠(2004)，「資料探勘在用藥安全上的應用：預測泛可黴素在腎衰竭病患上的用量適當性」，國立中山大學碩士論文。
3. 王秀雯(2004)，「應用資料挖掘技術於交通事故傷亡嚴重程度之研究」，國立嘉義大學碩士論文。
4. 王信勝(2001)，「整合分析層級程序與類神經網路之信用評分模型」，私立輔仁大學碩士論文
5. 林傑斌、劉明德(2002)，資料採掘與 OLAP 理論與實務，文魁資訊股份有限公司。
6. 林郁志(1998)，「都市地區肇事嚴重程度之分析研究」，國立成功大學交通管理科學研究所碩士論文。
7. 吳志文(2002)，「以類神經網路探討船舶事故之研究」，國立成功大學碩士論文。
8. 吳旭志、賴淑貞(2001)，Data Mining 資料採礦理論與實務，譯自 Michael J. A. Berry & Gordon S. Lionff。
9. 周雍傑(2000)，「以類神經網路探討都市地區肇事嚴重程度之研究」，國立成功大學碩士論文。
10. 郭一聰(2005)，「應用決策樹與類神經網路於應收帳款之呆帳預警模式研究」，私立中原大學碩士論文。
11. 陳文杰(2004)，「應用資料探勘技術於高速公路交通肇事次數之研究」，國立嘉義大學碩士論文。
12. 陳啟元(2003)，「資料探勘技術於健保資料之應用－以醫院門診服務點數預測為例」，國立中正大學碩士論文。
13. 陳俊傑(2002)，「應用資料探勘技術於道路交通事故資料分析之研究」，中央警察大學碩士論文。
14. 莊智仁(2003)，「應用個人違規記錄預測交通事故發生之研究」，國立嘉義大學碩士。

15. 黃昶斌(2004),「以類神經網路探討都市地區肇事嚴重程度」,國立交通大學碩士。
16. 葉怡成(1998),類神經網路模式應用與實作,儒林圖書公司。
17. 楊宗彥(2003),「運用類神經網路與決策樹技術預測股票報酬率」,私立逢甲大學碩士論文。
18. 楊憲國(2002),「以類神經網路推導路口交通事故成因與改善之研究」,私立逢甲大學碩士論文。
19. 彭慧雯(2001),「建構信用卡資料探勘架構及其實證研究」,國立台北科技大學碩士論文
20. 劉正揚(1997),「易肇事路口改善措施之研究」,國立交通大學交通運輸研究所碩士論文。
21. 劉秀鳳(2005),「應用類神經網路評估台 14 線公路邊坡崩壞潛能之研究」,朝陽科技大學碩士論文。
22. 謝森松(2004),「建立類神經網路模型預測住院跌倒之發生」,台北醫學大學碩士論文。
23. 魏建宏、李穎(2002),「國道客運班車旅行時間預測模式之研究」,中華民國運輸學會第 17 屆論文研討會中華民國 91 年 12 月。
24. 羅華強(2001),類神經網路—MATLAB 的應用,清蔚科技股份有限公司。

## 英文部分

25. Al-Ghamde, A. S., (2002). Using logistic regression to estimate the influence of accident factors on accident severity, Accident Analysis and Prevention, 34(6), 729-741.
26. Al-Ghamdi, A. S., (2003). Analysis of traffic accidents at urban intersections in Riyadh, Accident Analysis and Prevention, 35(5), 717-724.
27. Apte, C., Weiss, S., (1997). Data mining with decision trees and decision rules. Future Generation Computer System 13(2-3), 197-210.
28. Berry, M.J.A. and Linoff, G., (1997), Data Mining Technique for Marketing, Sale, and Customer Support, Wiley Computer.

29. Bevilacqua, M., Braglia, M., Montanari, R., (2003). The classification and regression tree approach to pump failure rate analysis, *Reliability Engineering and System Safety*, 79(1), 59-67.
30. Cabena P, Hadjinian P, Stadler R, Verhees J, Zanasi A. (1997), *Data Mining from Concept to Implementation*. Prentice-Hall : Englewood Cliffs, NJ.
31. Fayyad, U. M. Shapiro, G. P. and Uthurusamy, R. eds. (1996), *Advances in Knowledge Discovery and Data Mining*, Menlo Park, CA : AAAI/MIT Press.
32. Gupta, R. C. and Mann, L., (1975) "Effect of Certain Road Way Characteristics on Accident Rates for Two-Way Roads in Connecticut," *Transportation Research Record* 541, pp.50-54.
33. Hashemi, R.R., Le Blanc, L.A., Ruck, C.T., Shearry, A. ( 1995), *A Neural For Transportation Safety Modeling*.
34. Ivan, J. N. ( 1997 ) , "Neural Networks representations for arterial street incident detection data fusion," *Transportation Research Part C*, Vol. 5, No. 3/4, pp. 245-254.
35. Information Discovery, Inc. A characterization of data mining technologies and processes : An Information Discovery, Inc. White Paper. [http : //www.dmreview.com/](http://www.dmreview.com/)
36. Andrey, J. and Yagar, S, (1993) "A Temporal Analysis of Rain-Related Crash Risk," *Accident Analysis and Prevention*, Vol.25, No.4, pp.465-472.
37. Karlaftis, M. G., Golias, I., (2002). Effects of road geometry and traffic volumes on rural roadway accident rates. *Accident Analysis and Prevention*, 24(3), 357-365.
38. Kuhnert, P. M., Do, K. A., McClure, R., (2000). Combining non-parametric models with logistic regression : an application to motor vehicle injury data, *Computational Statistics & Data Analysis*, 34(3), 371-386.

39. Ogden, K.W. (1997) "The Effects of Paved Shoulders on Accidents on Rural Highways," *Accident Analysis and Prevention*, Vol.29, No.3, pp.353-362.
40. Marshall, R. J., (2001). The use of classification and regression trees in clinical epidemiology, *Journal of Clinical Epidemiology*, 54(6), 603-609.
41. Michalski, R.S., Bratko, I., Kubat, M., (1998). Machine learning and data mining methods and applications. John Wiley & Sons Ltd.
42. Mohamed Abdel-Aty, Anurag Pande, Identifying crash propensity using specific traffic speed conditions, *Journal of Safety Research* 36 (2005) 97– 108.
43. Moses, Leon N., and Savage, Ian (1994). The Effect of Firm Characteristics on Truck Accidents. *Accident Analysis and Prevention*, Vol.26, No.2, 173-179.
44. Mussone, L., Ferrari, A. and Oneta, M., (1999) "An analysis of urban collisions using an artificial intelligence model", *Accident Analysis and Prevention*, Vol. 31, pp.705-718.
45. Parker, D., West, R., Stradling, S. and Mantead, A.S.R., (1995a) "Behavioral characteristics and involvement in different types of traffic accident", *Accident Analysis and Prevention*, Vol. 27, No. 4, pp.571-581.
46. Parker, D., Reason, J.T. and Manstead, A.S.R., (1995b) "Driving negligences, diving violations and accident involvement", *Ergonomics*, Vol. 38, No. 5, pp.1036-1048.
47. Poch, M., Mannering, F., (1996). Negative binomial analysis of intersection-accident frequencies. *Journal of Transportation Engineering* 122(2), 105-113.
48. Subba Rao, P.V., Sikdar, P.K. and Krishna Rao, K.V., (1998) "Another insight into artificial Neural Networkss through behavioral analysis of access mode choice", *Computers, Environment and Urban Systems*, Vol. 22, No. 5, pp. 485-496.
49. Shankar, V., Mannering, F., Barfield W., (1995). Effect of roadway

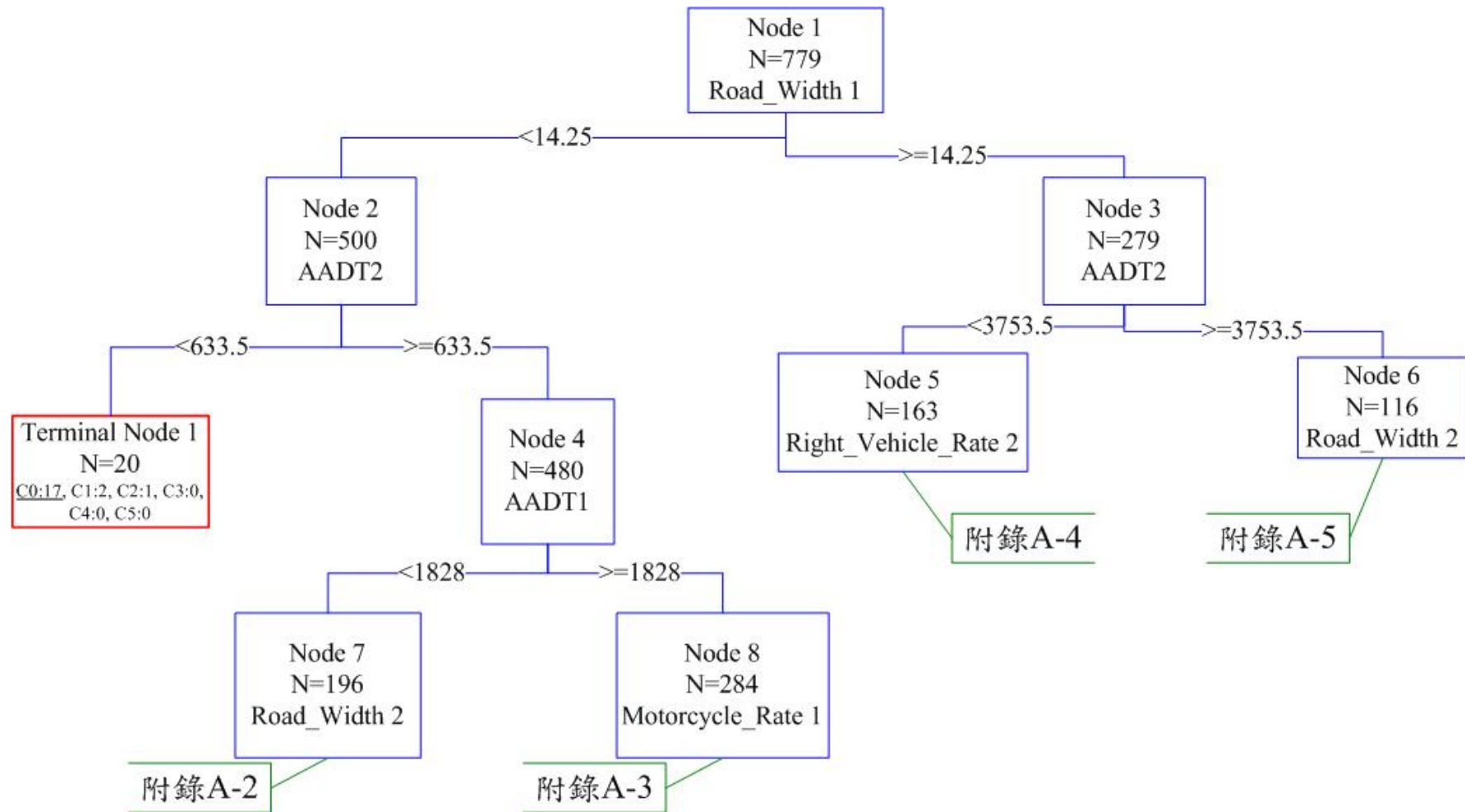
geometrics and environmental factors on rural freeway accident frequencies. *Accident Analysis and Prevention* 27(3), 371-389.

50. Shankar, V., Milton, J., and Mannering, F., (1997). Modeling accident frequencies as zero-altered probability processes : an empirical inquiry. *Accident Analysis and Prevention* 29(6), 829-837.
51. So Young Sohn, Sung Ho Lee (2003). Data fusion, ensemble and clustering to improve the classification accuracy for the severity of road traffic accidents in Korea, *Safety Science* 41, 1-14.
52. Zegeer, C. V., (1988). "Accident Effects of Sidelope and Other Roadside Features on Two-Lane Road," *Transportation Research Record* 1195.

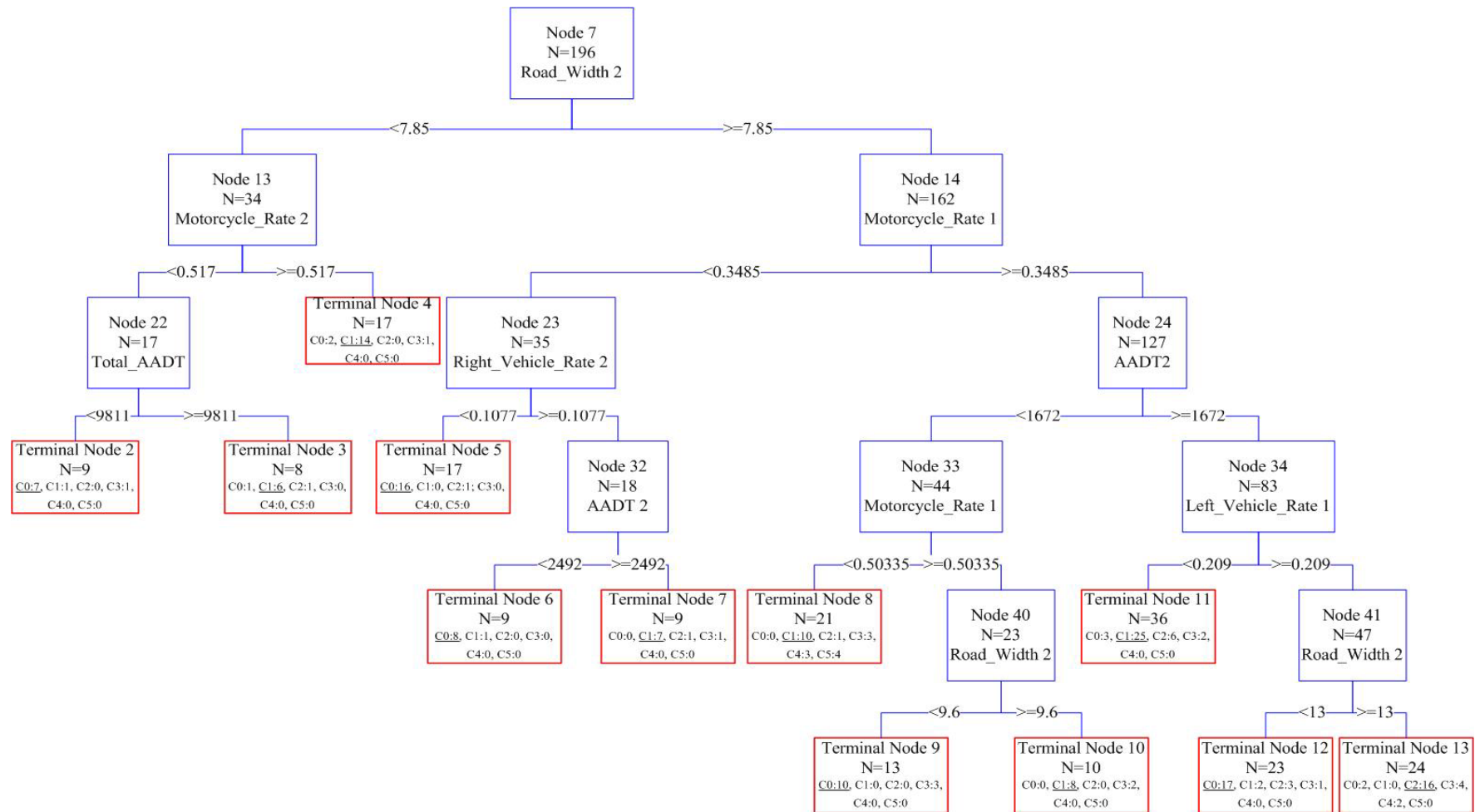




## 附錄 A Gini 決策樹

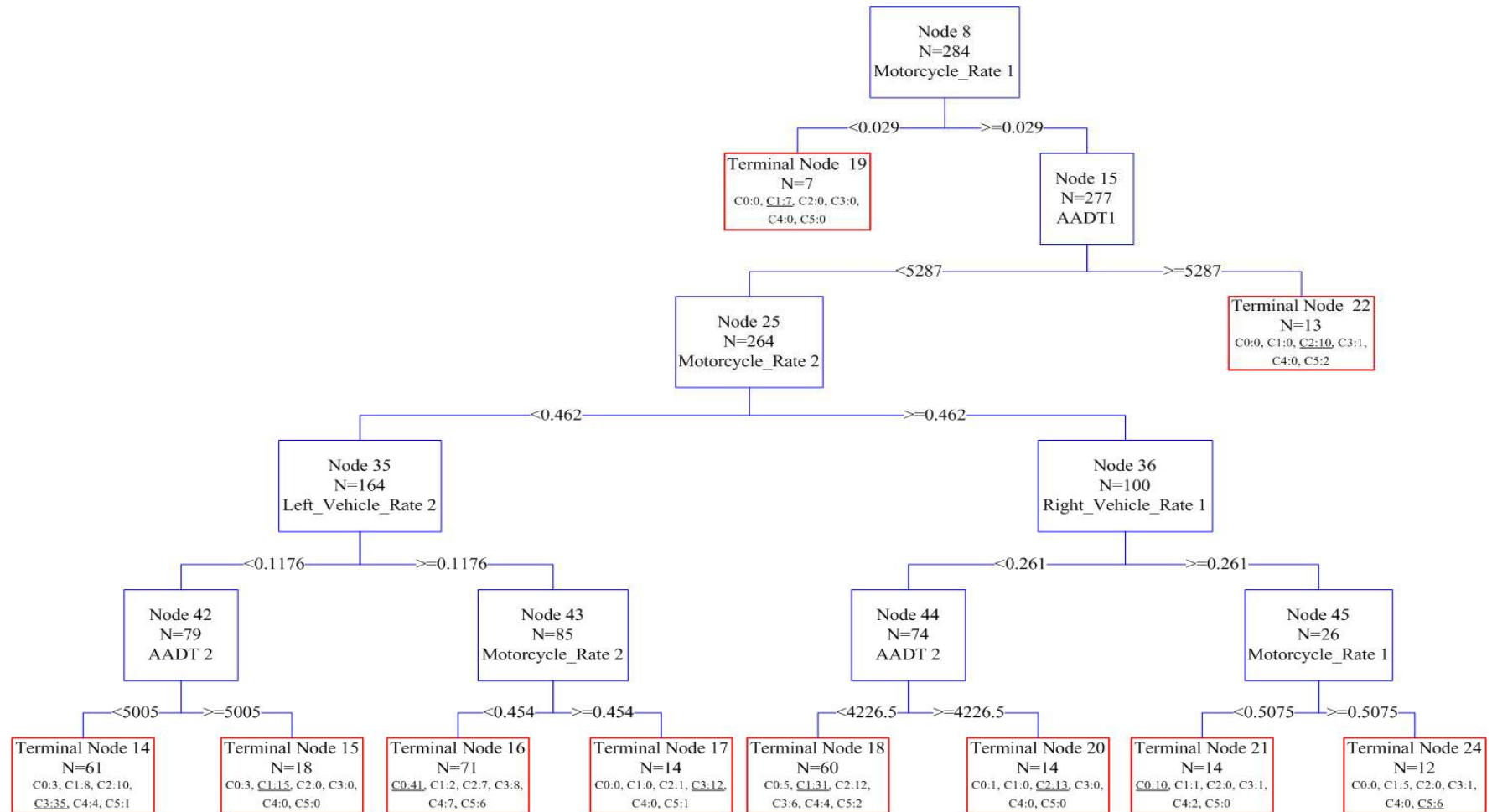


## 附錄 A Gini 決策樹

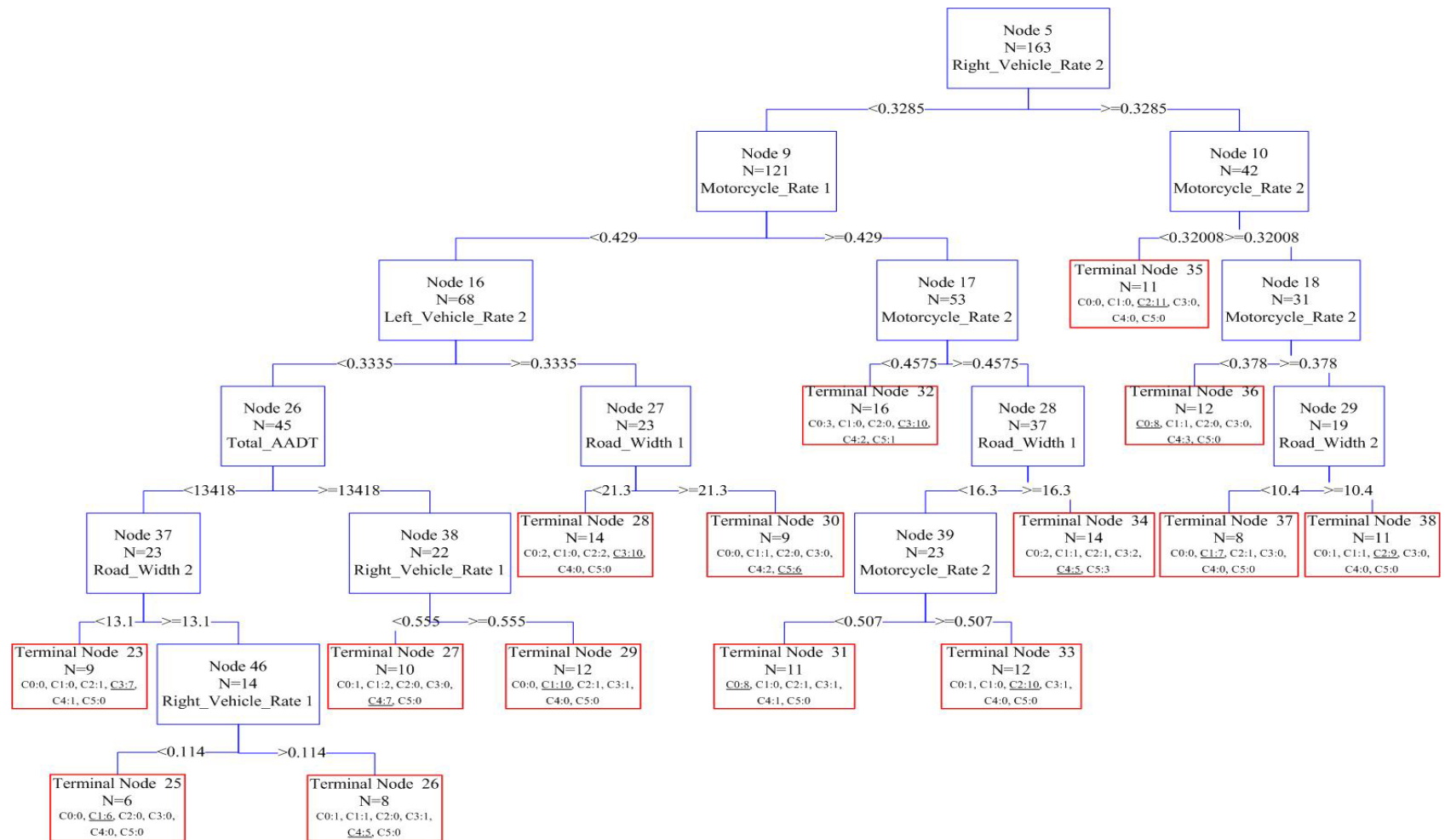




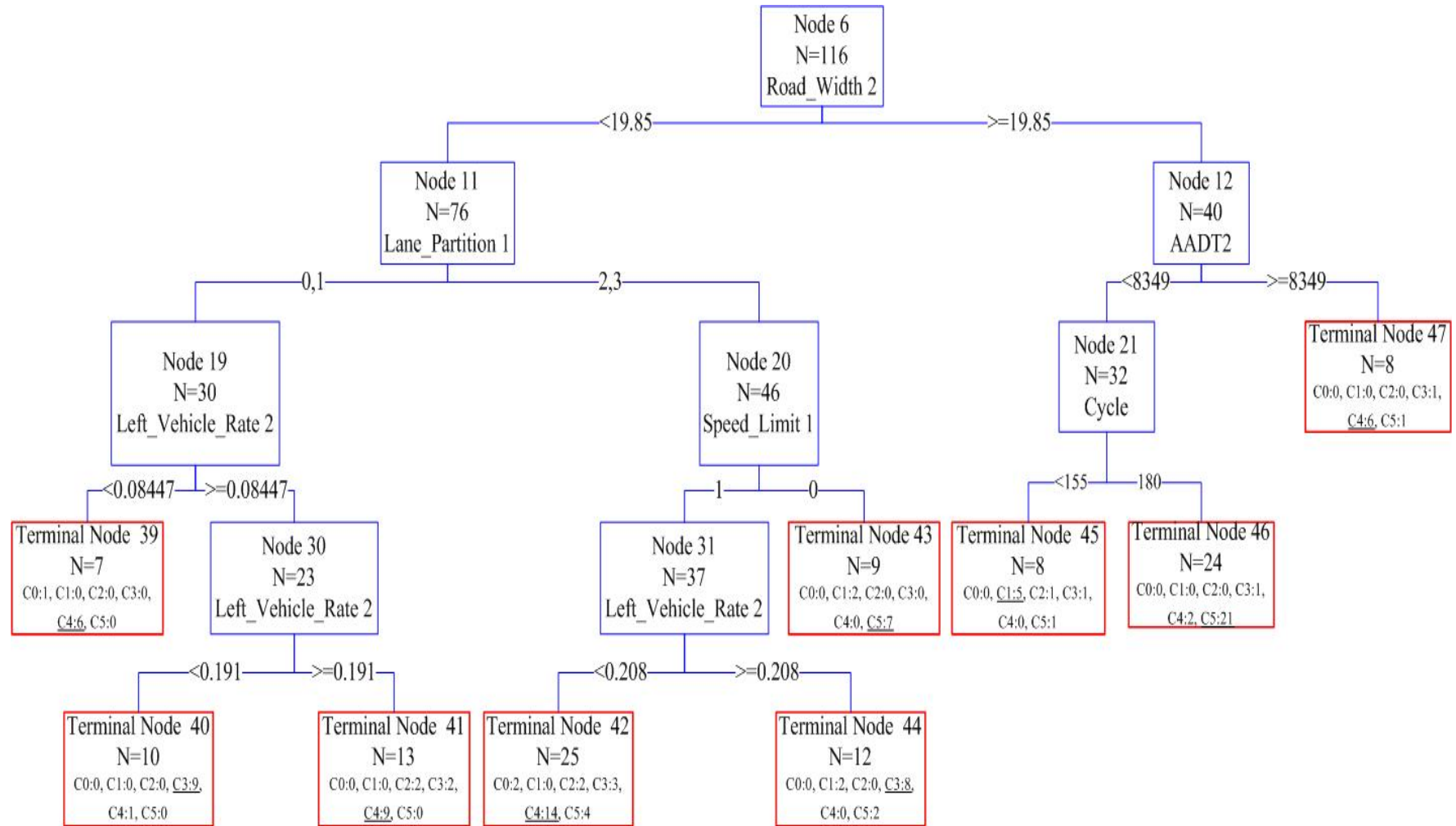
## 附錄 A Gini 決策樹



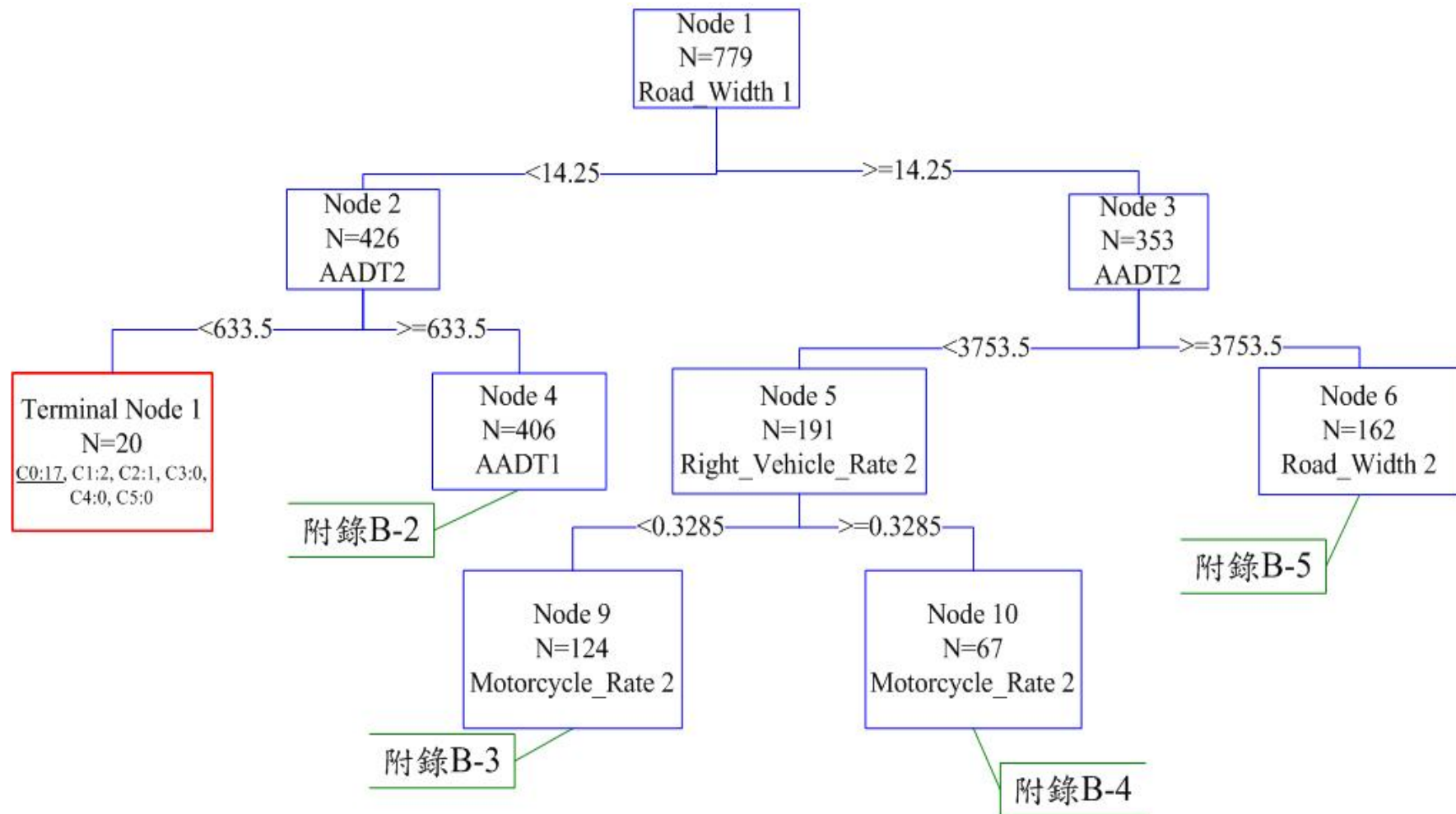
## 附錄 A Gini 決策樹



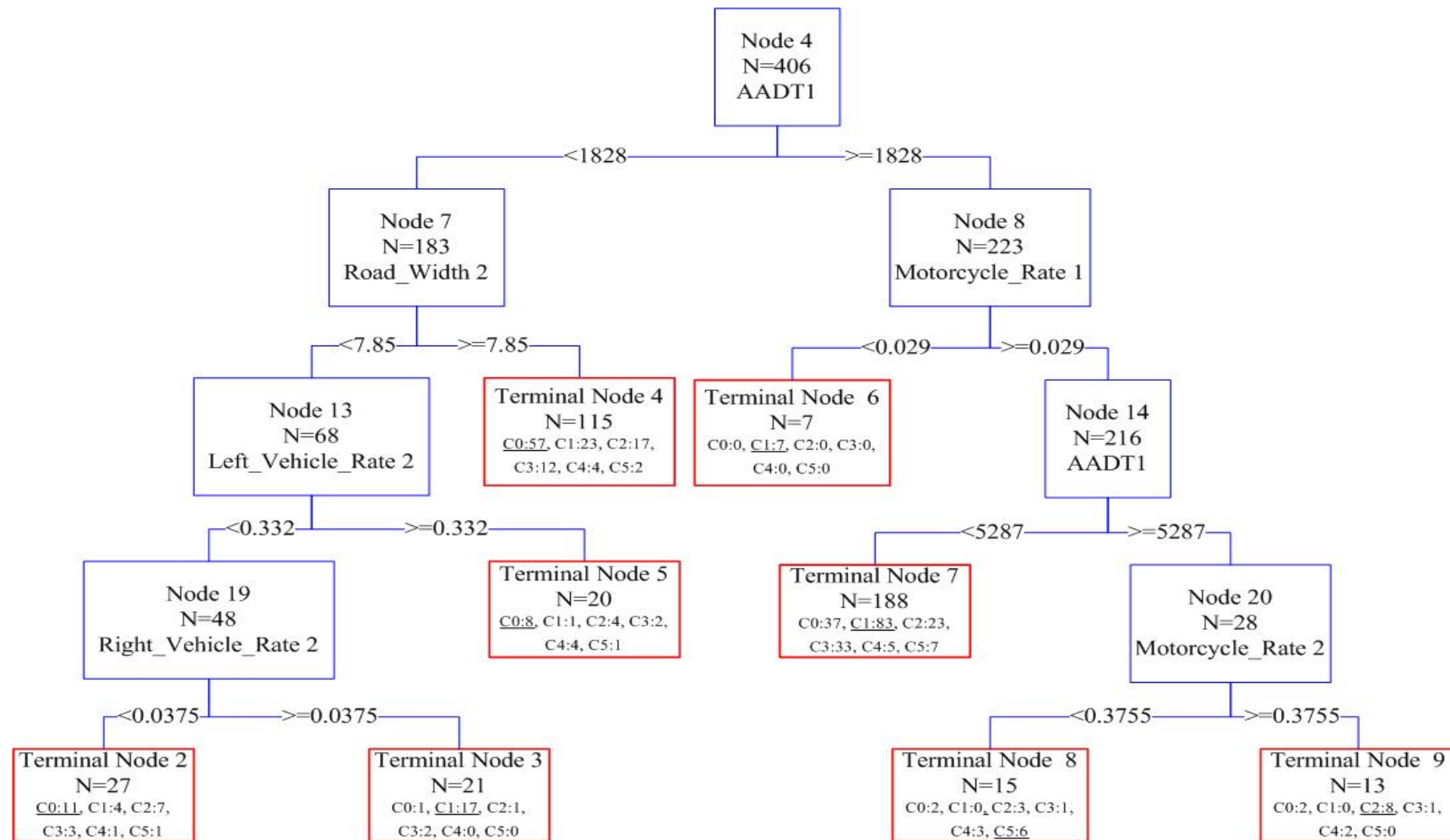
## 附錄 A Gini 決策樹



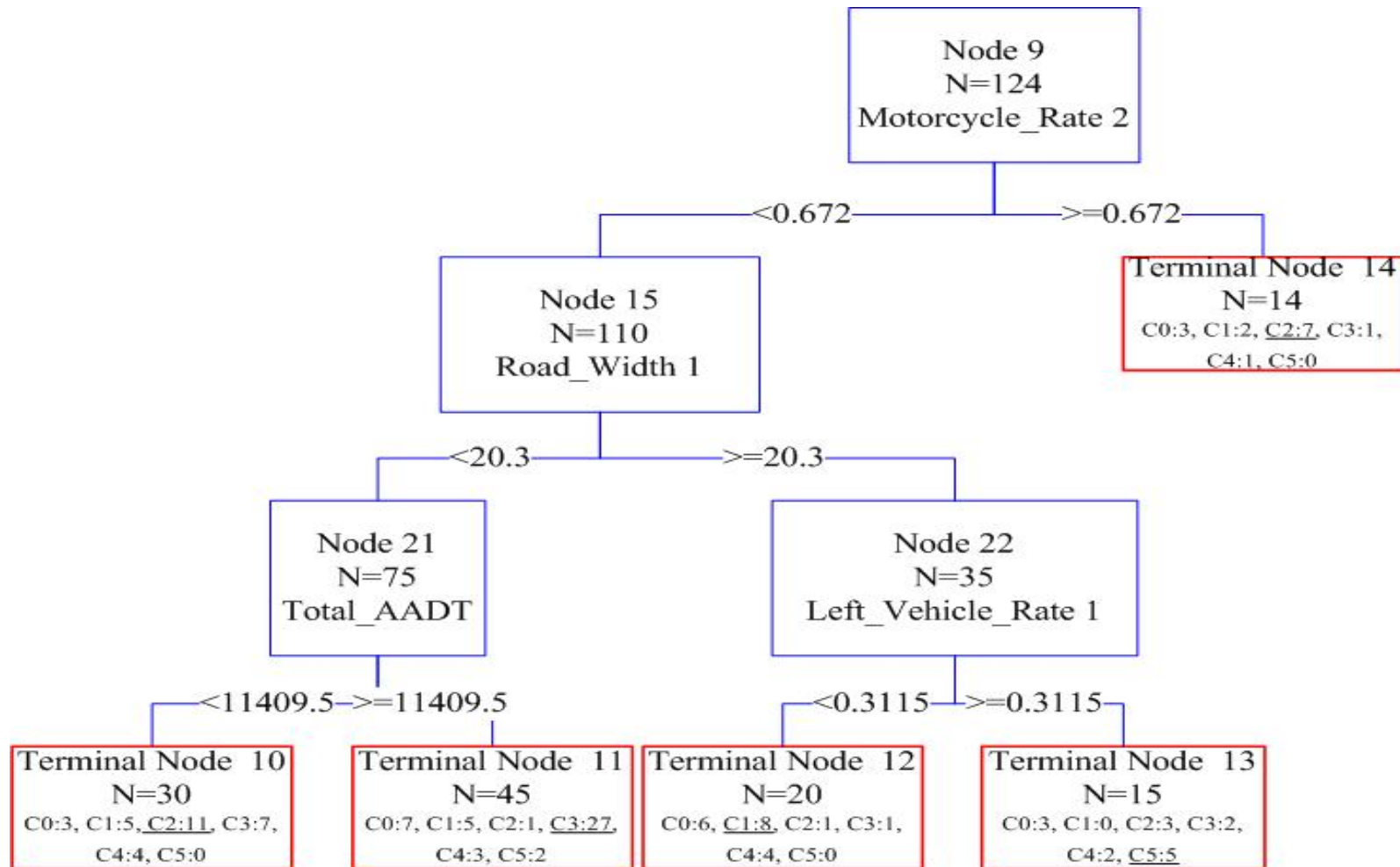
## 附錄 B Twoing 決策樹



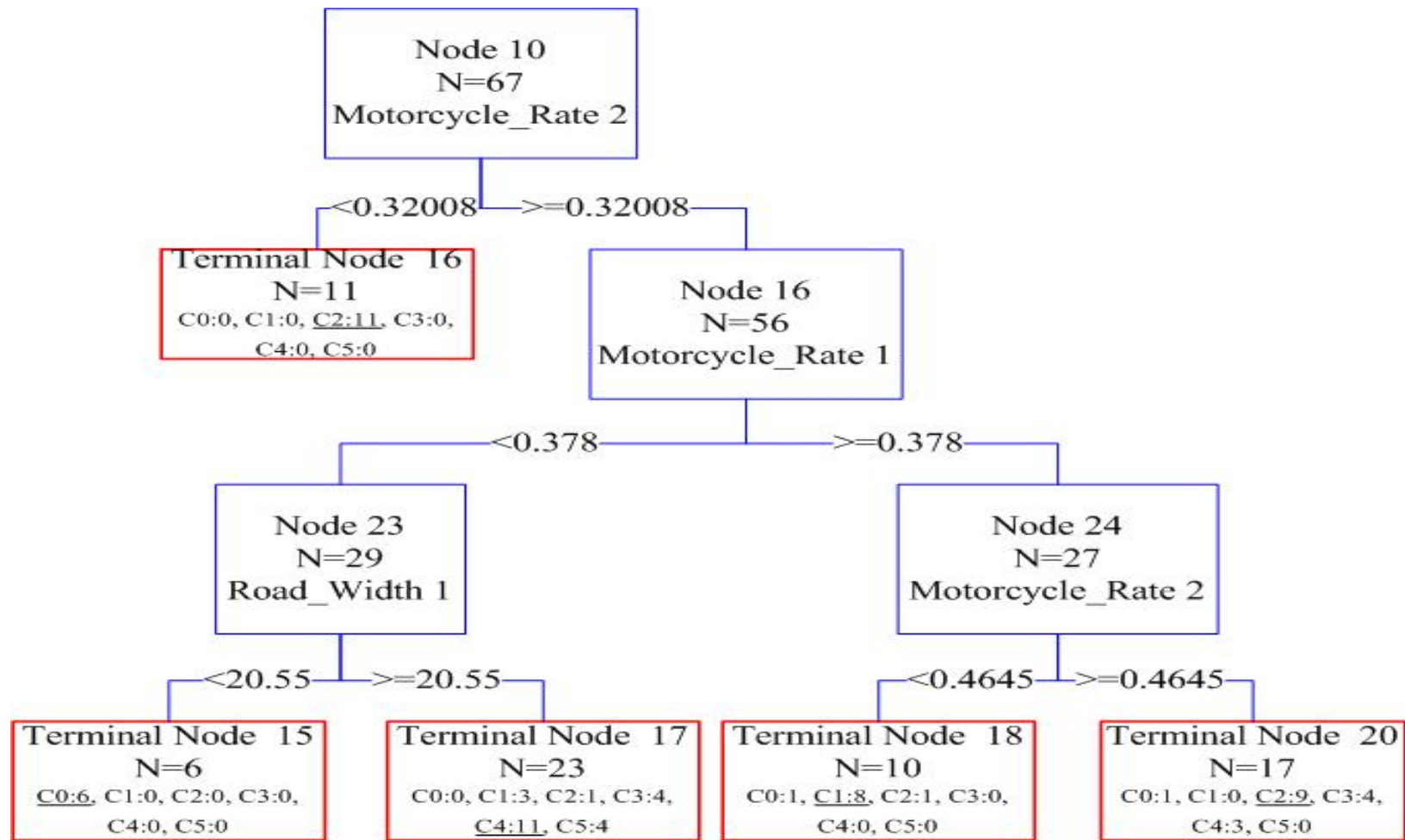
## 附錄 B Twoing 決策樹



## 附錄 B Twoing 決策樹



## 附錄 B Twoing 決策樹





## 附錄 B Twoing 決策樹

