

應用約略集合於高速公路肇事特性分析

洪境聰¹
許浚嘉²
王銘亨³
劉 霈⁴

摘 要

隨著車輛數量急遽成長，而台灣地區有限的土地資源與道路面積難以負載如此龐大的交通需求，造成許多地區道路交通秩序混亂，甚至引發交通事故造成人員與財務損失。依據高速公路 2014 年統計，該年度為 25,947.7 百萬車公里，每日平均約為 71.09 百萬車公里，因高速公路具有車流量大且車速快特性，故較於一般道路交通事故發生次數頻繁。有鑑於此，深入了解我國高速公路發生原因卻有研究之必要性。

肇事分析依其主題及目的可分為(1)肇事趨勢分析(2)肇事因果統計分析(3)危險路段鑒別三類，以往研究大多以「肇事因果統計分析」為主，其分析模式多採用統計分析如迴歸模式與判別分析模式，但這些模型均假設與預先定義應變數與自變數關係，因此當假設與現實因果關係有差異時將會造成可能性估計錯誤。近年來資料採礦技術逐漸成熟且在各領域應用皆有成效，在肇事因果分析曾應用關連規則與決策樹挖掘出新的因果關係與較佳的預測準確度。約略集合理論可在資訊不完整和資訊不一致下，用來約化資料集合、發掘隱藏的資料樣型和資料相關性，以產生有用的分類規則。故本研究以民國 100 年在高速公路路段所發生之 A1、A2 與 A3 類交通事故為範圍，推演肇事特性法則與肇事原因法則，經由約略集合分析所獲得法則具有高判中率，顯示約略集合理論應用於肇事分析，可以處理重複資訊、簡化屬性與挖掘屬性關係並發展適當之判定法則。

關鍵字：約略集合理論、高速公路肇事特性

一、緣 起

隨著近年來我國汽車持有數量不斷增加，道路交通事故數亦呈逐年遞增的趨勢。道路交通事故不僅造成社會經濟財產的損失，而且還危及人的生命。由於道路交通事故產生的原因可能涉及車輛零部件的機械故障、機動車駕駛員及行人對交通安全的重視程度和道路交通狀況等諸多因素的影響 (Golob, et al., 2004) (王秀雯, 2004)，故分析和研究道路交通事故與其相關影響因素之

¹開南大學運輸科技與管理學系副教授（聯絡地址：桃園市蘆竹區開南路 1 號，電話：03-3412500 轉 6083，E-mail：cthung@gapps.knu.edu.tw）。

²開南大學運輸科技與管理學系助理教授。

³臺灣警察專科學校交通管理科副教授。

⁴逢甲大學運輸科技與管理學系副教授。

間的關係，確定影響道路交通事故因素的主次因素，對於交通部門加強對道路安全的管理和對道路交通事故的控制具有重要的理論依據。

道路交通事故與相關因素的分析研究，通常有定性和定量研究兩種方式。定性研究，主要採用對影響其相關因素進行多方論證，但由於缺乏諸因素對其影響程度大小的必要研究，因此，無法清晰地瞭解影響其諸多因素中的主次之分。而定量研究是運用數學方法分析其諸多影響因素與道路交通事故的相關性，並依據所分析的結果，建立道路交通事故與影響其主要因素間的數學模型，對未來發展趨勢進行預測分析（羅均緯, 2005）(Quddus, et al., 2002)。但它通常要求具備大量完整的統計資料，且計算工作量較大。約略集合理論（Rough Set Theory, RST）是個新發展的數學工具，可作為資料庫資料處理、簡化並尋找適當決策的演算技術，並具備上述的系統化分析功能，RST 具備下列特性，使其自 1982 年已成功應用於許多不同領域中（Pawlak, 2001）(Pawlak, 1991)。

本研究利用約略集合處理不精確（imprecise）、不完整（incomplete）或不確定（uncertain）資料強有效的優點，以龐大的高速公路肇事資料庫為研究對象，應用系統化方法進行資料庫中肇事屬性表一致性、屬性重要性、屬性關聯性（association）、簡化肇事影響決策表等課題，由富含主觀與客觀特性的肇事資料庫中萃取有用資訊，增進肇事重要影響屬性的有效性。

二、資料探勘與交通事故

2.1 影響道路交通事故因素分析

從以往對於道路交通事故統計分析結果來看，道路交通事故的發生與車輛機械故障、非機動車和機動車輛駕駛者、行人對交通安全的重視程度以及道路交通狀況等諸多因素有極大的關連性。道路交通事故依其發生原因可分為四大類（周家慶 & 吳玉珍, 2003）：

- （一）人為因素：交通事故發生常與駕駛人之生理、心理狀態有關。生理方面包括駕駛人之官能反應能力、身體疲勞度、藥物或酒精作用等因素。在心理方面，則有心理感受之反應，如喜怒哀樂等情緒表現。另有原因如操作不當，或其駕駛技術或習慣等心理因素對交通事故的影響。
- （二）車輛因素：主要針對各車種之機件故障型態與肇事間之關係進行分析，如保養不良、煞車系統失靈、引擎系統、轉向系統、油電系統、爆胎等問題抑或裝載貨物方式欠妥而導致意外發生與交通事故發生之關聯性。
- （三）道路設施因素：道路線形之幾何條件，如坡度、轉彎半徑等；安全維護措施，如標誌、標線、反光導引器等；設計標準與施工品質，如該路段之設計速率、鋪面等，均會影響行車安全。
- （四）環境因素：本因素對駕駛人的影響諸如天候狀況，如雨、霧等，以及

道路之照明、行車速率、夜間眩光等，皆視為環境因素。

2.2 資料探勘用於交通事故之研究

過去交通事故的研究，大致上可以分為兩大類，一是以統計為主，另一則是資料探勘技術。統計方面的研究，有學者採用卜瓦松迴歸及負二項迴歸進行肇事因果分析 (林郁志, 1998)；也有學者使用依序羅機模式 (陳志和, 1999)及二元羅吉特模式 (楊思瑜, 2003)評估事故傷害程度。由於傳統的統計模型需要先行假設，如相依變數和解釋變數間呈現線性函數形式，所以統計模型雖早已被廣泛用於交通事故受傷嚴重程度之分析，但仍受到統計先天上之限制，當這些交通事故數據與統計假設不相符時，產生之統計模型將受到挑戰，甚至可以產生錯誤估計和推論 (Mussone, Ferrari, & Oneta, 1999)。

在交通事故分析上最常應用的探勘技術為分類法，利用分類及迴歸樹 (Classification and Regression Tree, CART)、類神經網路 (Artificial Neural Network, ANN)、決策數及支持向量機 (Support Vector Machine, SVM) 等多種模型以預測未來交通事故的發生趨勢，其研究結果也證明資料探勘技術對於現實交通事故資料有其適用性。但約略集合甚少應用於交通安全領域，故本研究將約略集合應用於照是特性、肇事原因、一般與重大事故之法則推演，探討不同肇事屬性與不同嚴重程度下，最適合應用何種法則推演，作為交通安全改善與注意方向。

三、約略集合理論介紹

真實世界中的資訊常是不精確 (imprecision)、不完整 (incomplete)、模糊 (vagueness)、不充分 (imperfection) 或不確定 (uncertainty) 的，為能獲得有用且有效的結果，我們必須具備能夠處理不確定與不完整資訊的能力。RST 是個新的數學方法，適於處理資料中不精確、模糊與不確定性的問題，其緣起於我們會針對有興趣的研究對象 (或稱目標、物件 object)，將許多相關的資訊、資料或知識相互連結 (associate) 起來，例如，若研究的對象是罹患某種疾病的病患，則疾病的徵狀就會形成許多與病患有關的資訊。若對象必須藉由相同或相似的資訊來辨別其特性時，便可能會產生相似 (similar) 或是難以辨識 (indiscernibility) 的情形，而此難以辨識關係的產生即為 RST 的數學理論基礎。

3.1 RST 背景與觀念

RST 是波蘭 Z. Pawlak 教授於 1982 年提出的一種智慧化決策分析工具 (Pawlak, 1981) (Pawlak, 1982) (Pawlak, 1985)，已被廣泛研究並應用於不精確、不確定、不完全訊息的分類分析和知識獲取。1991 年 Pawlak 教授出版了 Rough Set 專書，隨後連續召開 Rough Set 與知識發掘、Rough Set 與柔性計算等國際會議。RST 在醫療診斷、決策分析、機器學習、資訊檢索、近似推

理等領域均獲致優異的應用成果，成為知識發掘的重要工具。RST 中，知識是以對對象的“分類”能力為基礎，對象係指我們想像到的任何東西，包含真實的東西、狀態、抽象的概念、過程、時刻等。知識是由兩部份內容所構成：一是我們感興趣的領域中各種分類模式所構成的族群（family），其可提供明顯的事實；另一部份則是由明顯的知識衍生出不明確事實的推理能力。

我們將所有相類似對象的集合稱為基本要素（elementary），並依此形成知識的基本單元（atom），基本要素的任何聯集（union）可以是明確（crisp），即精確（precise）集合，否則就是約略（rough），即不精確或模糊的集合。依此定義，每個約略的集合均有邊界線（boundary-line）要素；亦即，除了明確集合明顯不存在邊界線要素外，其餘的集合皆無法將這些要素明確的分類為集合的某個隸屬（member）或其餘集（complement），亦即這些位於邊界上的案例無法將所獲得的知識作適當分類，因此，約略集合即可視為一種處理模糊觀念的數學模型。

RST 方法中，任何模糊的觀念可藉由一組概念上可視為精確的下界近似（lower approximation）與上界近似（upper approximation）來辨別。下界近似係由所有確定屬於該觀念的對象所組成；上界近似則由可能屬於該觀念的所有對象所組成，因此，“近似”即構成 RST 的兩個基本操作，如圖 1 所示。

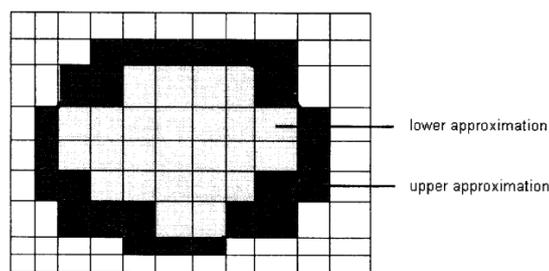


圖 1 下界近似與上界近似圖例

3.2 RST 與其他分析技術

模糊集合（Fuzzy Theory）與 RST 特別適合分析特殊的資料型態，尤其在處理不正確、不確定或模糊的知識時。RST 與模糊集合的基本觀念著重於不充分（imperfection）知識的兩個特點，即“難以辨識”（indiscernibility）與“模糊性”（vagueness）。模糊集合理論由 Zadeh 於 1965 年提出，已廣泛應用於鋪面及其他領域，RST 則由 Pawlak 於 1982 年提出，其應用雖在許多領域已非常盛行，但在肇事資料分析中則未多見。

一般集合理論中使用的是明確集合（crisp set），可以特定的要素定義之，也就是說，定義該集合時必須指出其要素為何。欲描述全域（universe）中各要素對該集合的歸屬度（belongingness）時，一般使用的是非 0 即 1 之隸屬函數（membership function），如圖 2(a)，其表示任何要素不是位於集合內就是集合外，而此並未考慮要素的不確定性。為處理不確定性的問題遂引入模糊集合的觀念，模糊集合係以一種封閉區間 $[0,1]$ 中之隸屬函數來定義，其允許要素的部份隸屬特性，如圖 2(b)，模糊性（fuzziness）係測度某事件發生

與否的程度。

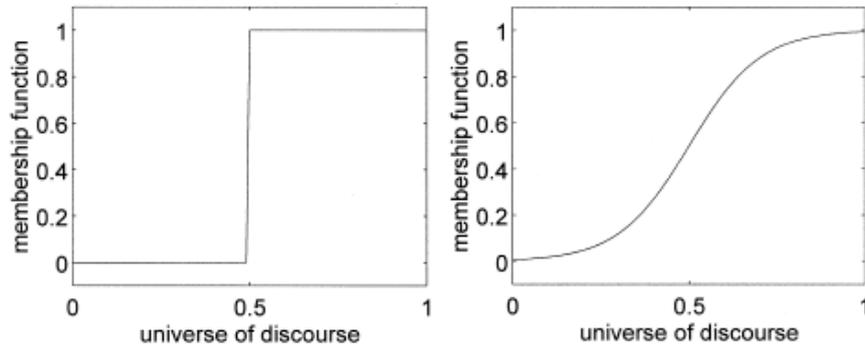


圖 2 (a)明確；(b)模糊集合理論之隸屬函數

然而在 RST 中，隸屬並非主要觀念，約略集合對模糊性與不確定性係以另一種不同的數學方法表示，RST 中對集合的定義與對全域中要素的資訊、知識與感知 (perception) 有關，亦即，RST 是依據可獲得且相關的資訊來總觀該全域的要素，因此兩種不同的要素依其可獲得之相關資訊判斷後，可能難以辨識導致看起來是一樣的。例如：兩個鋪面路段的 IRI 值分別為 1.33 與 1.86，就駕乘舒適程度而言均應視為非常平坦，但就該屬性 IRI 而言，兩者其實是難以辨識的，他們僅是約略集合中相較於“粗糙”或“中度”或其他與平坦度相關類別 (category)，而屬“平坦”的集合。

RST 對分類 (classification) 問題的解決能力係基於以下前提，降低資料中的精確程度使資料型態 (pattern) 更容易判斷，亦即 RST 可視為一種由不充分資料發掘事實的分析技術，其結果係以分類的形式，或由許多依據實例推導而得之決策法則來呈現，雖然 RST 的處理過程較之其他歸納式法則建立法 (inductive rule building methods) (例如，Id3 演算法) 不同，但對研究學者而言卻較容易理解。

除前述 RST 與模糊集合理論外，雲理論 (cloud theory) 與證據理論 (evidence theory) 亦可處理不確定性的問題，例如定性觀念與定量資料間的問題。事實上這 4 種理論間存在著特定關係，茲定義 x 為一個屬性， $u(x)$ 為針對某 X 等級 (class) 相應之隸屬值，分述如下：

- (一) RST 針對的是 x 的區間 $[u_{\min}(x), u_{\max}(x)]$ ，一個屬性有許多相對應的值，即一對多，最後的決策為該屬性 “is”、“is not” 或 “is maybe” 某等級。對集合 X 而言，介於下界近似與上界近似間的資料是約略的，不確定是否屬於集合 X 。RST 擴展明確集合的使用，著重於因為不完整、不充分或無法完整取得資訊導致之不確定性，較之其他方法，RST 可更緊密且確實的描述存在於真實世界中的實際對象，包含確定的與不確定的。
- (二) 模糊集合針對的是一個值 $u(x)$ ，其相對於 x 的模糊隸屬函數 $u(x)$ ，該關係表示一個屬性對應的一個函數值。模糊集合也是明確集合的延伸，可對不確定性問題進行分類，其著重的是因模糊、朦朧或不明瞭資訊導致之不確定性，但模糊隸屬函數的決定較為困難，且模糊集合倚賴主觀經驗，一旦模糊隸屬度 $u(x)$ 決定後便缺少不確定的特性。

- (三) 雲理論具備三種數值特性，在 x 範圍內以 $u(x)$ 值定義間斷 (discrete) 資料， $(x, u(x))$ 稱為雲滴 (cloud drop)，間斷程度可由隸屬值 $u(x)$ 決定， $u(x)$ 的範圍與區間則不確定。雲模型也可呈現介於定性觀念的語意詞彙 (linguistic term) 與其數值間的不確定特性。
- (四) 證據理論，也名為 Dempster-Shafer 理論、證據的 Dempster-Shafer 理論、或簡易函數 (belief function) 的 Dempster-Shafer 理論，其使用簡易函數與似真 (plausibility) 函數，與 RST 的上界與下界近似相似。整體的明確性 (crispness) 測度可解釋為由 Dempster-Shafer 邏輯觀點所獲得之簡易值 (belief value)，唯簡易函數完全倚賴經驗。由於證據理論與 RST 間的相似性，故目前仍持續針對兩個理論進行探究。

綜觀前述分析，雖然存在許多技術可用來處理不確定性資料類型的問題，然而 RST 技術的優勢在於其數學理論的簡單與其驚人有效的處理問題能力。RST 係針對不準確資料進行推理或發掘這些資料間的關係，因此其研究與統計學非常相關，但卻是完全不同的方法。相較於其他方法，使用 RST 有以下優點：

- (一) 評估對象與決策間關係之特定屬性重要性
- (二) 移除所有重複的對象與決策，以獲致最少的屬性次集合，對於決策所
- (三) 進行的分類仍可獲得滿意的近似結果。
- (四) 針對特定的決策類別產生最具代表性的推演模式。

近年來約略集合進行分析的研究對象愈見廣泛，因其理論特性不需服從任何假設條件，讓應用更具有彈性，同時是用於質化與量化的變數。該理論經由數據挖掘隱藏於其中的現象，並且符合自然與法方式來展現這些決策法則並剔除多餘或無用的訊息，讓其決策法則更易理解，使用者不須對最終模型的參數進行解讀。例如張簡曉馨利用某民間發卡銀行的資料建立信用卡呆卡辨識模式 (張簡曉馨, 2005)；張家瑞將約略集合應用於鋪面管理當中建立鋪面破損原因推測法則 (張家瑞, 等, 2005)。

四、以 RST 建立肇事原因法則

本研究資料來自於國道公路警察局登陸國道高速公路所有事故資料作為分析資料庫，其調查表格中在民國 100 年度造成 A1、A2 與 A3 事故共計 17319 筆，資料型態為文字和數字，表格欄位包括天候、道路類別、路面狀況等資訊。當獲得研究資料後，先進行資料一致化整理，剔除異常樣本及極端值。另經文獻整理得知肇事是由車、路、人與環境四大構面造成，本研究依據交通事故資料表選取 24 個屬性，將此 24 個屬性依其性質與定義歸納於車輛、道路與環境屬性類別 (因獲得之資料庫無駕駛相關資訊，故未分析當事人因素)，以約略集合進行法則推演，分析不同肇事分類下，三大屬性類別嚴重程度高低。

表 1 屬性三大類別

屬性類別	屬性
車輛屬性	車種
道路屬性	所在縣市 所在鄉鎮 國道編號 事故發生方向 道路類別 速限 道路型態 事故位置 路面鋪裝 路面狀態 路面缺陷 車道劃分設施 分道設施—快車道 分道設施—快慢車道 路面邊緣
環境屬性	時 天候狀況 光線 障礙物 視距 號誌種類 號誌動作

本研究採用 RSES(Rough Set Exploration System)約略集合軟體進行實證資料分析，RSES 具備強大資料庫分析能力，可輕易進行移除重複記錄、簡化屬性、挖掘屬性關性、推演法則等，對於肇事應用上自動化的知識探索與萃取具有效率。約略集合理論將 23 種肇事特性進行刪減分析，刪減後為 20 種肇事特性如表 3 所示，最小屬性與核心屬性為 20 種肇事特性，共刪減掉發生縣市、號誌動作/號誌種類、路面邊緣共 3 種肇事特性。當正區域(Pos. Reg.)

=1，表示刪減後的資料表與原始資料表的嚴重程度是一致的，本研究肇事特性正區域達 0.9711；刪減的穩定係數(Stability Coefficient, SC)代表動態折減的穩定性，當值為 1 時最穩定。

表 2 肇事特性折減

1-2	肇事特性數量	正區域	穩定係數	折減
1	20	0.9711	1	時,區,道路,方向,天候,光線,道路類別,速限,道路型態,事故位置,路面鋪裝,路面狀態,路面缺陷,障礙物,視距,號誌種類,分向設施,快車道或一般車道間,快慢車道間,車種
2	20	0.9711	1	時,區,道路,方向,天候,光線,道路類別,速限,道路型態,事故位置,路面鋪裝,路面狀態,路面缺陷,障礙物,視距,號誌動作,分向設施,快車道或一般車道間,快慢車道間,車種

以 17319 筆記錄推演肇事原因法則，推演出 14130 條法則，有以下三種形式：各組肇事特性對性一種嚴重程度、各組肇事特性對應二種嚴重程度、各組肇事特性對性三種嚴重程度。各組肇事特性對應一種嚴重程度有 13980 法則符合、各組肇事特性對應二種嚴重程度有 149 法則符合、各組肇事特性對應三種嚴重程度有 1 法則符合。

表 3 各組肇事特性對應法則數量

肇事嚴重程度	法則數量
1	57
2	914
3	13009
1、2	1
1、3	3
2、3	145
1、2、3	1

將折減成 20 種肇事特性後產生 14130 條法則，對於 A1 類別其判中率為 100%、A2 類別其判中率為 99.1%、A3 類別其判中率為 93.6%，其整體判中率為 93.9%。對於 A3 類別的影響因素較為分散，故對於其判中率較其他類別為低。

表 4 驗證 14130 條肇事特性法則結果

		推 演 肇 事 嚴 重 程 度			
實 際 肇 事 嚴 重 程 度		A1	A2	A3	判中率
	A1	62	0	0	1
	A2	3	1055	7	0.991
	A3	19	1019	15154	0.936
	正確推演率	0.74	0.51	1	

五、結論與建議

RST 是波蘭大學 Z. Pawlak 教授於 1982 年提出的一種智慧化資料決策分析工具，已廣泛研究並應用於不精確、不確定、不完全的訊息分類分析及知識獲取。RST 為肇事資料的知識發掘開啟了一條新途徑，其分析結果顯示在共同肇事屬性數量下，嚴重程度多果會影響整體判中率高低，其中 A1 類別的法則判中率為 100%。RST 可進一步簡化該一般性資料並推演最佳化決策演算法，使其能在具備一般性資料內涵的前提下進行最佳化知識的萃取。

參考文獻

- 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. 嘉義: 嘉義大學運輸與物流工程研究所碩士論文.
- 周家慶, & 吳玉珍. (2003). 永續運輸資訊系統—交通事故資料分析資料研究. 台北: 交通部運輸研究所.
- 林郁志. (1998). 都市地區肇事嚴重程度之分析研究--以台南市為例. 台南: 國立成功大學交通管理研究所碩士論文.
- 張家瑞, & 洪境聰. (2005). 以約略集合理論推演鋪面養護策略之研究. 行政院國家科學委員會專題研究計畫書成果報告.
- 張簡曉馨. (2005). 信用卡呆卡式別模式之建立. 台南: 南臺科技大學國際企業研究所碩士論文.
- 陳志和. (1999). 都市地區肇事嚴重程度預測模式之研究. 台南: 國立成功大學交通管理研究所碩士論文.
- 楊思瑜. (2003). 小型車事故特性分析及嚴重程度預測模式之研究--以桃竹苗

- 地區為例。台中：逢甲大學交通工程與管理研究所碩士論文。
- 羅均緯。(2005)。應用約略集合理論預測交通事務嚴重程度之研究。台南：南台科技大學國際企業研究所碩士論文。
- Golob, T., Recker, W., & Alvarez, V. (2004). Freeway Safety at a Function of Traffic flow. *Journal of Accident Analysis & Prevention*, 36(5), pp. 933-946.
- Mussone, L., Ferrari, A., & Oneta, M. (1999). An analysis of urban collisions using an artificial intelligence model. *Accident Analysis and Prevention*, 31(6), pp. 705-718.
- Pawlak, Z. (1981). Information systems-theoretical foundations. *Information Systems*, 6, pp. 205-218.
- Pawlak, z. (1982). Rough Sets. *International Journal of computer and information Sciences*, 11(5), pp. 341-356.
- Pawlak, Z. (1985). Rough sets and Fuzzy sets. *Fuzzy Sets and system*, 17, pp. 99-102.
- Pawlak, Z. (1991). *Rough sets: Theoretical Aspects of Reasoning About Data*. London: Kluwer Academic Publishers.
- Pawlak, Z. (2001). A primer in rough sets: a new approach to drawing conclusions from data. *Cardozo Law Review*, 22, pp. 1407-1415.
- Quddus, M., Noland, R., & Chin, H. (2002). An analysis of motorcycle injury and vehicle damage severity using ordered probit model. *Journal of safety research*, 33(4), pp. 445-462.