

住宅竊盜案隱性鏈結犯罪關聯模式之研究

林建隆
刑事警察局研發室警務正
Lin.rufus@gmail.com

王朝煌
中央警察大學資訊管理系教授兼主任
jwang@mail.cpu.edu.tw

摘要

當刑事案件發生後，承辦員警須填寫刑案紀錄表，內容包含犯罪成因、犯罪習癖、準備措施、犯罪方法、犯罪工具五項犯罪模式變數資料，及刑案發生地點、日期／時間等案件屬性資料及嫌犯特徵等相關資料。本研究擬藉由分析住宅竊盜案犯罪紀錄，計算犯罪成因、犯罪習癖、準備措施、犯罪方法、及犯罪工具等變數的資訊熵值與機率分佈，並歸納犯案區域特性，建立案件間或與嫌犯間的隱性鏈結，發展犯罪關聯模式。藉以作為研判住宅竊盜案是否為同一犯罪嫌疑人／集團連續犯案之參考。並以蒐集實務機關住宅竊盜案犯罪紀錄，運用實驗評估犯罪關聯模式之準確度、召回率及整體綜效。

關鍵字: 隱性鏈結分析，犯罪關聯模式，犯罪模式，犯案區域特性，犯案時段。

一、緒論

鏈結分析乃犯罪偵查的重要工具之一。社會網絡如血緣關係、財務交易、及通聯紀錄等，常被用來建立犯罪個體(entities)間的鏈結[16,20]。然而運用資料探勘技術分析資料庫的內容而建構的隱性鏈結，則鮮少用於建立犯罪個體間的鏈結。雖然犯罪模式、犯案時段、及犯案區域特性等常被用於犯罪剖繪且極具運用價值，但其運用效能常取決於剖繪專家的經驗智慧。因此研擬資料探勘技術，分析刑案資料，挖掘犯罪模式、犯案時段、及犯案區域特性之隱性鏈結，以輔助犯罪偵查工作之進行，實為迫切之課題。

依路卡交換原理，犯罪者進入犯罪現場之每一接觸，除了在現場留下跡證之外，也會帶走一些物證[18]。犯罪剖繪專家經由分析犯罪現場跡證的種類、數量、位置、及狀態，可據以重建犯罪過程[9]及推導嫌疑犯的犯罪模式，進而縮小偵查範圍[14]。犯罪剖繪主要基於犯罪模式或犯罪手法[14,19]，如被害人的共同性特徵、犯罪使用的武器、犯罪的敵意、有無凌虐及性侵害等跡證，據以推導犯罪者的人格特質、性別、年齡、生活背景、及可能的物理特徵如身高體重等等，並依據嫌疑犯與剖繪特徵的契合程度，比對評估可能的罪犯。

犯罪乃社會變遷，如經濟情況、社會文化等各種因素交互影響的結果。例如遷徙活動較少及社會較為穩定的地區犯罪率通常較低、地域性城市的暴力犯罪與竊盜犯的增加率可能較高、而大都會地區毒品濫用的程度可能較市郊嚴重[6]。因此地域特性形成重要的犯罪因素，地域特性可分為「自相關性」與「負的空間自相關」，自相關性指鄰近地

區事件有較大之相互影響作用，舉凡鄰近地區其相關性較大，而與遠離地區其相關性則較少之情形，此情形稱為正的空間自相關；反之若與鄰近地區差異性較大而與遠離的地方差異性較小，則稱為負的空間自相關[3]。地點與時間對犯罪分析同樣具重要性，如前所述每個人的一天都是在空間與時間的路徑中，相關實證研究結果也顯示：犯罪案件間的關係強度與案發地點間隔的距離成反比關係[1]。

時間與地點對犯罪分析同樣具重要性，因為每個人的一天都是在時間與空間的路徑當中[7]。例如闖空門的竊盜案多發生於白天的上班時間，而搶劫落單上班族的搶案多發生於傍晚。換句話說，犯罪時間也是犯罪的重要因素之一。以「小時」為變數的犯罪分析變化是非常大的，可以很多種分類，例如一星期有168小時。以一天24小時來劃分，首先要決定的是一天的起始點，學者Marcus Felson與Erika Poulsen提出以5:00AM作為一天的起始，因為夜生活的人在清晨5點大多已經返家，而早起的人也大多在5點開始活動[5]。因此一天是由5:00AM開始，到隔天的4:59結束。同時利用中位數及第1、3四分位數作為犯罪分析的指標，利用這些指標分析犯罪的集中與離散程度。例如分析紐約搶奪案與臺北市搶奪與住宅竊盜案之犯罪時間指標，紐約的搶奪案件集中在晚上至凌晨，特別是20:30至00:35分，亦即有四分之一的案件集中在這4小時5分之內。相較於臺北市的搶奪案或住宅竊盜案，則較集中於白天，臺北市搶奪案有四分之一的案件集中於12:35至16:50這4小時15分鐘，住宅竊盜案則集中於11:00至15:10的4小時10分鐘，詳如表一：

表一. 紐約與臺北市犯罪時間指標分析表

案由與地區 時間指標	紐約 搶奪案	臺北市 搶奪案	臺北市 住宅竊盜
起始	05:00	05:00	05:00
第1四分位	15:00	12:35	11:00
中位數	20:30	16:50	15:10
第3四分位	00:35	20:50	21:00

在已破或是有目擊者的案件，犯罪嫌疑犯、被害人、以及犯罪的鏈結是具體而明確的，但是未破或缺乏目擊者的案件，犯罪現場的證據便成為推導犯罪鏈結最重要的憑據。刑案發生後，承辦員警必須填寫刑案紀錄表，內容可區分為「案件基本資料」及「犯罪模式(modus operandi, MO)」。基本資料如案件類別、案發地點、案發時間、嫌犯及被害人資料等等；犯罪模式包含犯罪成因(crime cause, CC)、犯罪習癖(crime habit, CH)、準備措施(preparation action, PA)、犯罪方法(crime method, CM)、及犯罪工具(crime tool, CT)五項變數，且每一變數均有預設值可供挑選。

我國警察機關自1970年起開始使用電腦處理資料，目前已累積大量的犯罪資料。近年來由於資訊科技的進步，資料分析技術已有長足的進展，尤其資料探勘 (data mining) 技術已日趨成熟。本研究擬藉由分析住宅竊盜案犯罪紀錄，計算犯罪成因、犯罪習癖、準備措施、犯罪方法、及犯罪工具等變數的資訊熵值與機率分佈，並歸納犯案時段與區域特性，藉以建立案件間或與嫌犯間的隱性鏈結，發展犯罪關聯模式。本文結構為：第

二節探討相關文獻，第三節說明住宅竊盜犯罪關聯模式之建構，第四節描述實驗與報告評估結果，第五節為本文結論。

二、文獻探討

本研究運用的理論及技術主要包含犯罪模式、鏈結分析、及資訊檢索的向量空間模型，分述如下：

1. 犯罪模式

「犯罪模式」原出於拉丁文的Modus Operandi，英譯為Methods of Operation，日本譯為「犯罪手口」，國內有人譯稱為「犯罪方式」或謂「犯罪手法」，意指慣行犯於作案時所慣用的犯罪手段、方法或型式。根據奧地利刑事法學者漢斯克魯茲的研究結果，犯罪模式具必存性、反覆性(慣行性)、及固定性之特性。犯罪模式主要應用於連續犯、常業犯、及職業犯犯罪行為的特徵分析，由於犯罪者的理性選擇，如安全考量、容易得手及實質利益等因素，往往會反覆運用自己最得意的手段與方法從事犯罪行為[10,19]，因此可從其作案的手段、方法或形式分析歸納建檔[2,15]，作為偵辦刑案之參考。此外犯罪人口可以劃分成大部份的偶發性犯罪人(occasional offender)及少部份的習慣犯罪者(chronic offender)。但那少部份的習慣犯罪者卻犯了相當大比例的犯罪案件。因此如將犯罪者之獨特性格與行為特質，歸納分析其犯罪模式建立資料庫，當犯罪發生後，經由分析案件之犯罪模式與特徵，即可據以比對搜尋資料庫，找出可能的涉嫌者或疑似犯罪集團之連續犯行。

2. 鏈結分析

鏈結分析(link analysis)乃用以處理及分析個體間關係的技術，由人與人、物與物或人與物間關係發展出來的應用，其理論基礎主要是來自於數學的分支，稱為圖形理論(graph theory) [4,11,16,20]。鏈結分析及其視覺化技術，通常將具有關聯性的資料以圖形(graph)展示，以視覺化方式展示其組織與特殊型態，以利進一步分析，找出隱藏其中的資訊。由於各種關係中，隱含了許多豐富的網絡資訊，例如：透過電話通聯分析，可以推論用戶為一般住民或業務推銷員，以及用戶的夥伴關係；分析醫生的處方籤可以推知其特殊藥商關係，另外分析醫病關係，可發現兩位醫生是否共同治療一位病人或曾經治療過同一位病人。鏈結分析可將這些隱藏的網絡關係挖掘出來。美國聯邦調查局即利用鏈結分析結合不同地方的犯罪情報，協助解決犯罪問題。

3. 向量空間模型

資訊檢索領域的向量空間模型[12]，主要用於衡量文件查詢與文件間的接近程度，乃搜尋引擎的核心技術之一。首先以自動索引技術計算關鍵詞集合 $\mathbf{K} = \{k_1, k_2, k_3, \dots, k_M\}$ 及其與文件集合 $\mathbf{D} = \{d_1, d_2, d_3, \dots, d_N\}$ 之每一文件的關聯，求得下列關聯係數矩陣：

$$\begin{array}{cccccc}
 & d_1 & d_2 & d_3 & \dots & d_N \\
 k_1 & a_{11} & a_{12} & a_{13} & \dots & a_{1N} \\
 k_2 & a_{21} & a_{22} & a_{23} & \dots & a_{2N} \\
 k_3 & a_{31} & a_{32} & a_{33} & \dots & a_{3N} \\
 \dots & \dots & \dots & \dots & \dots & \dots \\
 k_M & a_{M1} & a_{M2} & a_{M3} & \dots & a_{MN}
 \end{array}$$

a_{ij} 為關鍵詞 k_i 與文件 d_j 的關聯，乃以詞頻/反文件詞頻(term frequency / inverse document frequency, TF/IDF)公式計算：

$$a_{ij} = tf_{ij} * \log \frac{N}{Cf_i} \quad (1)$$

其中 tf_{ij} 為關鍵詞 k_i 在文件 d_j 出現的頻率， N 為文件總數， Cf_i 為含關鍵詞 k_i 的文件個數。意乃關鍵詞 k_i 與文件 d_j 的關聯與關鍵詞 k_i 在文件 d_j 出現的頻率成正比，與含關鍵詞 k_i 的文件數則成反比關係。

將關鍵詞與文件的關聯係數矩陣轉置，文件 d_j 可以 $(a_{j1}, a_{j2}, \dots, a_{jM})$ 表示，類似地查詢 q_k 也可以 $(a_{k1}, a_{k2}, \dots, a_{kM})$ 表示。假設關鍵詞間相互獨立，文件 d_j 與查詢 q_k 的關聯可以向量內積(一般稱為向量空間模型)求得，公式如下：

$$\text{association}(d_j, q_k) = \sum_{i=1}^M a_{ji} * a_{ki} \quad (2)$$

詞頻/反文件詞頻的計算方式及向量空間模型，已廣應用於資訊檢索(information retrieval)及文字探勘(text mining)領域。

4. 資訊熵值

在資訊理論中，熵被用來衡量一個隨機變數出現的期望值。1948年，克勞德·艾爾伍德·夏農將熱力學的熵，引入到資訊理論，因此它又被稱為夏農熵(Shannon entropy)。

資訊熵也稱為信源熵、平均自資訊量[13,19]。一個 X 值域為 $\{x_1, x_2, x_3, \dots, x_n\}$ 的隨機變數熵 H 定義為： $H(X) = E(I(X))$ ， E 代表了期望值函數，而 $I(X)$ 是 X 的資訊量， $I(X)$ 本身為隨機變數。以一個二進位的編碼系統而言，如果 P 代表 X 的機率分布(probability distribution)， $P = \{p(x_1), p(x_2), p(x_3), \dots, p(x_n)\}$ ，則 X 的平均訊息量熵 $H(X)$ 可以下列公式計算求得：

$$H(X) = \sum_{i=1}^n p(x_i) I(x_i) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i) \quad \text{其中}$$

$$I(x_i) = - \log_2 p(x_i), \text{熵 } H(X) \text{ 的單位為位元(bit)}. \quad (3)$$

假設26個英文字母(letter, L)在文章中出現次數平均的話，每個字母的平均訊息量(編碼或傳輸所需位元數)為： $I(L) = - \log_2 \frac{1}{26} = 4.7$ 位元。另外如果文章使用3000個常用漢字(character, C)，且每個常用漢字在文章中出現頻率相同，則每個常用漢字的平均資訊量為： $I(C) = - \log_2 \frac{1}{3000} = 11.6$ 位元。

¹ 註：配合關聯係數矩陣轉置，式中 a_{ij} 表示文件 d_i 與關鍵詞 k_j 的關聯。

如果每個字母或常用漢字出現的機率不一，則每個字母或常用漢字的編碼或傳輸所需位元之平均數，可以上述計算 $H(X)$ 之運算式求得。例如一篇文章含3000個英文字母的文章中，字母'A'出現2975次，而'B'~'Z'等25個字母各出現一次，則每個字母的平均訊息量為：

$$- \left(\frac{2975}{3000} * \log_2 \frac{2975}{3000} + 25 * \frac{1}{3000} * \log_2 \frac{1}{3000} \right) = 0.11$$

從上例中可知，字母出現頻率越不平均，每個字母的平均訊息量越低。另外從中英文文章中字母及漢字的平均訊息量比較，可知隨機變數值或事件項目越多，則每個隨機變數值或事件的平均資訊量也越大。綜合而言，隨機變數或事件項目越多、越平均分布，則每個隨機變數或事件的平均資訊量越大。由於平均訊息量的計算公式因為和熱力學中熵的玻爾茲曼公式一樣，所以也被稱為熵(entropy)。

三、犯罪關聯模式

本研究研擬的住宅竊盜犯罪關聯模式包含犯罪模式、犯案時段、及犯案區域特性三項資訊，犯罪模式包含犯罪成因、犯罪習癖、準備措施、犯罪方法、及犯罪工具等五項變數，每一變數的權重擬以資訊熵值代表其權重。此外，由於每一變數各有其值域及機率分佈，本研究擬以每一變數值之機率分佈作為每個值權重的計算基礎。結合變數的權重與值的權重，計算犯罪模式的關聯係數。犯案時段與犯案區域特性則以實際案例統計結果，研擬趨近之數學關聯模式。未來將延伸犯罪關聯模式以納入更多的犯案特性，如犯案季節及被害人特徵等資訊，力求周延。

1. 犯罪模式關聯

(1) 變數權重

由於隨機變數或事件項目越多、越平均分布，則每個隨機變數或事件的平均資訊量越大，代表變數值越具區辨力。因此在發展關聯模式時，熵值大的變數，權重配置應越高，反之則反。犯罪模式的每一變數皆有其值域與機率分佈，變數的資訊熵值皆可由分析住宅竊盜案犯罪紀錄加以計算。例如，竊盜犯罪習癖的統計分佈與資訊熵值如表二：

經統計住宅竊盜案犯罪模式變數熵值平均為：犯罪原因1.04，犯罪習癖1.49，準備措施1.56，犯罪方法2.28，犯罪工具1.89。

(2) 值的權重

根據犯罪學的研究指出：少部份的習慣犯罪者犯了相當大比例的犯罪案件[7]，從實際案例資料分析也發現，習慣犯罪者所犯案件的平均數遠大於偶發性犯罪人。經進一步分析，也發現習慣犯罪者雖非以同一犯罪模式重複犯案，但其犯罪模式變數的值域遠小於所有案件的犯罪模式變數。換言之，習慣犯罪者的犯罪模式變數有其獨特的機率分佈。本研究根據資訊檢索領域“詞頻/反文件詞頻”的概念[12]，分別計算每一習慣犯犯罪模式變數值的機率分佈，整合所有案件犯罪模式變數的機率分佈，作為個別犯罪模式變數值的權重。習慣犯罪者P的犯罪模式變數F之值V的權重為：

$$\log \frac{P_{V_{PF}}}{P_{V_F}} \quad (4)$$

其中 $P_{V_{PF}}$ 為值V 在習慣犯罪者P的犯罪模式變數 F 出現的機率，而 P_{V_F} 為值V 在所有案件犯罪模式變數 F 出現的機率。換言之，犯罪模式變數值在一嫌犯所犯的案件出現頻

率越高，表示該嫌犯常用該技倆或具備該習性，越能代表該嫌犯。然而若該犯罪模式變數值也常出現於其它案件，即其它嫌犯也具備該習性或慣用該技倆，則該犯罪模式變數值作為嫌犯關聯的效應相對調降。

表二. 住宅竊盜案犯罪習癖變數值分布與熵值

代碼	犯罪習癖	次數	$p(x_i)$	$-\log_2 p(x_i)$	$p(x_i) * -\log_2 p(x_i)$
null	未填代碼	3852	0.70770	0.50	0.35
07	趁人不備或熟睡或入浴	600	0.11023	3.18	0.35
08	闖空門(趁人不在、無人)	419	0.07698	3.70	0.28
03	專偷住宅	300	0.05512	4.18	0.23
37	無犯罪習癖	220	0.04042	4.63	0.19
36	其他習癖	23	0.00423	7.89	0.03
06	專偷特定物品(如女人用品)	8	0.00147	9.41	0.01
15	自行攜帶工具作案	7	0.00129	9.60	0.01
05	專偷辦公室含企業機關學校	4	0.00073	10.41	0.01
24	調戲或強暴婦女或性虐待	2	0.00037	11.41	0.00
28	故意破壞現場	2	0.00037	11.41	0.00
02	專偷工廠倉庫	2	0.00037	11.41	0.00
01	專偷商店	1	0.00018	12.41	0.00
29	帶面具手套作案	1	0.00018	12.41	0.00
32	作案前喝酒或服用禁藥	1	0.00018	12.41	0.00
13	遺留攜帶物	1	0.00018	12.41	0.00
合計		5443		熵值	1.49

(3) 犯罪模式之關聯係數

類似於資訊檢索領域的向量空間模型，假設犯罪模式變數成相互獨立的關係，結合變數權重與值的權重，習慣犯罪者P與犯罪案件Q在犯罪模式項目的關聯係數值可以下列公式求得：

$$\begin{aligned}
 & E(CC) \times match(P_{CC}, Q_{CC}) + \\
 & E(CH) \times match(P_{CH}, Q_{CH}) + \\
 & E(PA) \times match(P_{PA}, Q_{PA}) + \\
 & E(CM) \times match(P_{CM}, Q_{CM}) + \\
 & E(CT) \times match(P_{CT}, Q_{CT})
 \end{aligned} \tag{5}$$

其中E(CC)為犯罪成因(crime cause)變數資訊熵之平均值，

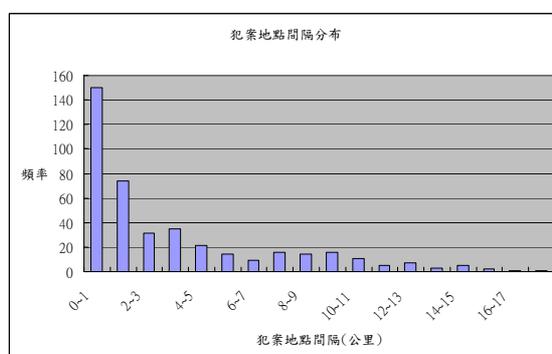
$match (P_{cc} , Q_{cc}) = \log \frac{P_{V_{PCC}}}{P_{V_{cc}}}$, 若 $P_{cc} = Q_{cc}$ 且 $\log \frac{P_{V_{PCC}}}{P_{V_{cc}}}$ 大於0, 否則設為0, 依此類推。

2. 住宅竊盜案犯罪區域關聯

經統計住宅竊盜慣犯之犯案地點間隔的頻率統計結果，其中36.14%的案件間隔在1公里以內，1公里以上則據減。住宅竊盜習慣犯連續犯案地點間隔如表三及圖一所示。

表三. 住宅竊盜連續犯相鄰犯案地點間隔

距離(公里)	次數	百分比
0~1	150	36.14%
1~2	74	17.83%
2~3	31	7.47%
3~4	35	8.43%
4~5	21	5.06%
5~6	14	3.37%
6~7	9	2.17%
7~8	16	3.86%
8~9	14	3.37%
9~10	16	3.86%
10~11	11	2.65%
11~12	5	1.20%
12~13	7	1.69%
13~14	3	0.72%
14~15	5	1.20%
15~16	2	0.48%
16~17	1	0.24%
17~18	1	0.24%

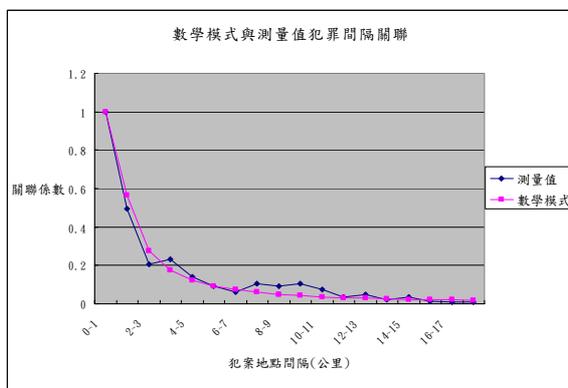


圖一: 住宅竊盜連續犯相鄰犯案地點間隔分佈圖

住宅竊盜案之犯罪距離間隔呈現等比級數下降之趨勢，惟1至2公里則較0~1公里劇減，因此，數學關聯模式以1公里以下，關聯係數為1，1公里以上則為距離的多次方倒數。住宅竊盜案之犯罪區域特性之權重值計算數學函式如下：

$$\text{關聯係數} = \begin{cases} 1, & \text{犯案地點間隔在1公里以內} \\ \frac{1}{d^{1.4}}, & \text{犯案地點間隔大於2公里} \end{cases}$$

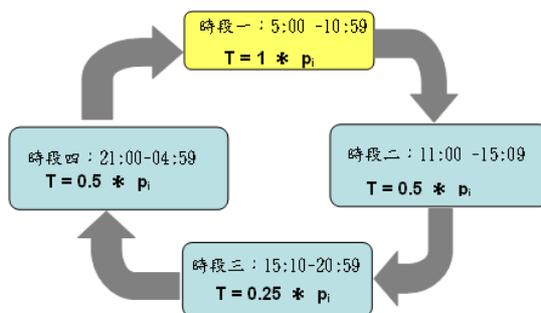
經計算趨近實際頻率分布之次方數，以1.4次方可得最少之平方差和(為0.0240)。比較犯罪距離頻率分布之正規化數值，與本研究研擬之數學關聯係數如圖二：



圖二: 住宅竊盜案數學模式與測量值犯罪間隔關聯近似圖

3. 犯罪時段關聯

本研究以案件在24小時的分布作為時間切割單位，並以5:00 AM作為一天的起始。每一天係由5:00 AM開始，至隔天4:59 AM結束，再以住宅竊盜案發生時間之分布，切割為中位數，第1、3四分位數，將24小時分為四個時段。本研究住宅竊盜案5,443件依發生時間分布的四分位數為5:00，11:00，15:10及21:00。本研究以連續犯模型的犯罪時間在四個時段的機率作為時段之權重(pi)。時段關聯之計算，則以時段相同之權重為1，次鄰時段之權重為1/2，再次鄰時段權重為1/4，以住宅竊盜案之犯罪時間關聯係數計算為例，如下圖三所示：



圖三: 住宅竊盜案犯罪時段關聯係數計算

四、住宅竊盜實驗結果與分析

1. 實驗設計

(1) 犯罪關聯模式設計

本研究研擬的犯罪關聯模式包含「犯罪模式(Profile, P)」、「犯罪區域特性(Distance, D)」及「犯罪時間(Time, T)」三項變數，三項變數均以min-max normalization方式標準化為0-1之數值。住宅竊盜案關聯計算分別以單變數，雙變數組合及三變數組合計算關聯係數，其中雙變數組合及三變數組合以1, 2, 4等比例做為變數的計算權重。單變數之計算，分別以P, D, T計算個別關聯係數，共計3組。雙變數以兩兩組合方式計算(PD, PT, DT)，並以1,2,4之權重比例計算，共計15組。三變數則將PDT以1, 2, 4等權重比例組合，例如：P1D2T4，P4D1T1...等，扣除重複之比例，如P1D1T1與P2D2T2及P4D4T4等三組均為1:1:1之權重比例，共計19組。

(2) 評估方法

本研究以北部某警察局轄內5443已破之竊盜案件，分別計算每一竊盜案與9個住宅竊盜連續模型之關聯係數。並分以0.9，0.8，0.7及0.6等作為關聯門檻值，計算9個住宅竊盜模型之準確率(Precision, P)、召回率(Recall, R)及綜效(F-Measure, F)。準確率係指：判斷為與某一住宅竊盜模型關聯的案件，實際為該嫌犯所犯案件之比例。召回率係指：判斷為與某一住宅竊盜模型關聯的案件且實際為該嫌犯所犯之案件數與該嫌犯所犯案件總數之比值。綜效F則為“ $2 * R * P / (R + P)$ ”。例如嫌犯A犯案總數為10件，假設在關聯門檻值0.8時，20個案件判斷為與嫌犯A關聯。但這20個案件只有8個案件屬於嫌犯A所犯的案件，則準確率 $P = 8/20 = 0.4$ ，召回率 $R = 8/10 = 0.8$ ， $F = 2 * 0.4 * 0.8 / (0.4 + 0.8) = 0.53$ 。

2. 實驗結果

(1) 單變數關聯模式

以「犯罪模式」、「犯罪區域特性」及「犯罪時間」等單一變數關聯模式，計算5443竊盜案與9個住宅竊盜連續模型之關聯。其準確率、檢出率及綜效之平均值如表四、表五及表六所示：

表四.犯罪模式關聯模式之實驗結果

犯罪模式(Profile, P)				
關聯門檻	0.9	0.8	0.7	0.6
評估指標				
準確率	0.75	0.73	0.61	0.48
檢出率	0.13	0.18	0.24	0.31
綜效	0.22	0.28	0.34	0.38

表五. 犯罪區域特性關聯模式之實驗結果

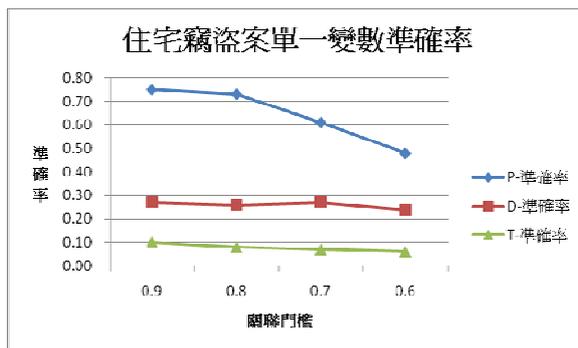
犯罪區域特性(Distance, D)				
關聯門檻	0.9	0.8	0.7	0.6
評估指標				
準確率	0.27	0.26	0.27	0.24
檢出率	0.36	0.38	0.42	0.46
綜效	0.31	0.31	0.33	0.31

表六.犯罪時間關聯模式之實驗結果

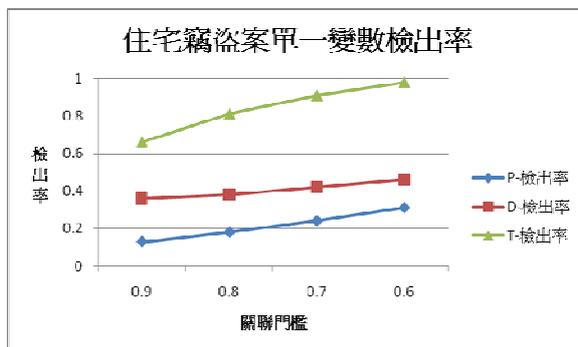
犯罪時間(Time, T)				
關聯門檻	0.9	0.8	0.7	0.6
評估指標				
準確率	0.10	0.08	0.07	0.06
檢出率	0.66	0.81	0.91	0.98
綜效	0.17	0.14	0.13	0.11

由表四、表五及表六比較分析，犯罪模式關聯模式，在關聯門檻0.9時，有較佳的準確率0.75，且在關聯門檻0.6時，有較佳的綜效0.38；犯罪時間關聯模式，在關聯門檻0.6時，有較佳的檢出率0.98。

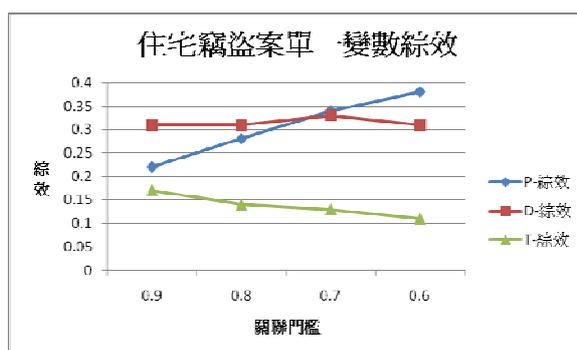
在單一變數的關聯模式中，準確率以犯罪模式效果最好，而且隨著門檻值降低時，準確率有明顯下降趨勢，如圖四所示。而檢出率則以犯罪時間效果最好，在關聯門檻0.9時，檢出率高達0.66，且隨著關聯門檻降低，有明顯的上升趨勢，在關聯門檻0.6時，檢出率更高達0.98，而犯罪模式及犯罪區域特性亦隨著關聯門檻降低而上升情形，但是，並無非常顯著上升之走勢，如圖五所示。在綜效方面，關聯門檻在0.9及0.8時以犯罪區域特性效果較好，而關聯門檻在0.7及0.6則以犯罪模式效果最好且隨著關聯門檻降低而有顯著上升之趨勢，如圖六所示。



圖四.單一變數準確率比較圖



圖五.單一變數檢出率比較圖



圖六.單一變數綜效比較圖

(2) 變數兩兩組合關聯模式

雙變數P, D, T兩兩以1:1、1:2、及1:4權重比例組合後,計算5443竊盜案與9個住宅竊盜連續模型之關聯。實驗結果以犯罪模式與犯罪區域特性權重比例為4:1 (即P4D1)組合,在關聯門檻為0.6,所得之綜效0.41為最佳,其準確率、檢出率及綜效平均值如表七所示:

表七.住宅竊盜案P4D1實驗結果

犯罪模式與犯罪區域特性(P4D1)				
關聯門檻	0.9	0.8	0.7	0.6
評估指標				
準確率	0.56	0.73	0.67	0.52
檢出率	0.03	0.13	0.22	0.33
綜效	0.07	0.21	0.33	0.41

(3) 三變數組合關聯模式

三變數P, D, T以1, 2, 4權重比例之各種組合後,計算5443竊盜案與9個住宅竊盜連續模型之關聯。實驗結果以P1D2T4組合在關聯門檻0.7,所得綜效0.46為最佳,其準確率、檢出率及綜效平均值如表八所示

表八.住宅竊盜案P1D2T4實驗結果

P1D2T4				
關聯門檻	0.9	0.8	0.7	0.6
評估指標				
準確率	0.51	0.43	0.35	0.15
檢出率	0.07	0.28	0.66	0.84
綜效	0.13	0.34	0.46	0.25

3. 結果分析與討論

分析表四~八的實驗結果,單一變數以犯罪模式的關聯可得最佳的綜效0.38,犯案區域特性次之0.33,犯案時段再次之0.17。主要原因為住宅竊盜案之犯罪時間通常無法確知,僅能依被害人「發現時間」判斷,因此犯罪時間通常為「發現時間」,而非正確

的犯罪時間²。雙變數兩兩組合，犯罪區域特性與犯罪時間以4:1(P4D1)之權重比例，最佳綜效落在關聯門檻值0.6，準確率0.52及檢出率0.33，可得最佳綜效0.41。結合犯罪模式、犯案區域特性、及犯案時段等三項資訊，以1:2:4(P1D2T4)之權重比例，最佳綜效落在關聯門檻值0.7，準確率0.35及檢出率0.66，可得最佳綜效0.46，即以本研究所蒐集分析的資料實驗，在最佳權重比例的條件下，在準確率平均為35%的情形下，平均可以找出66%嫌犯所犯的案件。此外，單一變數的最佳綜效為0.38，雙變數兩兩組合的最佳綜效為0.41，而三變數組合的最佳綜效為0.46，也印證「掌握相關資訊越多，越有利決策」的經驗智慧。

五、結論

本文提出犯罪隱性鏈結的概念與定義，並以統計及機率分佈為基礎，建立犯罪模式及犯案區域特性等隱性鏈結，發展犯罪關聯模型以計算嫌犯與住宅竊盜案間之關聯係數。本研究並蒐集實務機關之住宅竊盜案犯罪紀錄，以實驗評估犯罪關聯模式之效度。實驗結果顯示，本文所提出的犯罪關聯模式可以找出住宅竊盜案關聯性，最佳綜效落在關聯門檻值0.7，準確率0.35及檢出率0.66，可得最佳綜效0.46。即在準確率平均為35%的情形下，可以找出66%嫌犯所犯的案件。在本研究所蒐集分析的三個資訊項目中，以犯罪模式的關聯可得較佳的綜效，犯案區域特性次之，犯案時段再次之。此外，三變數組合的最佳綜效大於變數兩兩組合的最佳綜效，變數兩兩組合的最佳綜效又大於單一變數的最佳綜效，印證「掌握相關資訊越多，越有利決策」的經驗智慧。

未來研究方向，將納入更多的犯案特性，如犯案季節及被害人特性等資訊，力求周延強化犯罪關聯模型。另外對於半結構化或非結構化的文件資料，亦將發展分析技術，期能加以整合分析，以蒐集更廣泛的相關資訊，進一步提昇犯罪關聯模型的效度。

參考文獻

- [1] 陳仁智，地緣剖繪技術應用於連續街頭強盜搶奪犯罪偵查之研究，*中央警察大學刑事警察研究所碩士論文*，2004。
- [2] 黃富源、范國勇、張平吾，*犯罪學概論*，三民書局，2005。
- [3] 廖友祿、吳國清、王朝煌、黃俊能，「*犯罪資料庫管理系統*」應用技術服務委託專案研究報告，臺北市政府警察局，2010。
- [4] M. J. Berry, and G. Linoff, *Data Mining Techniques: For Marketing, Sales, and Customer Support*, John Wiley & Sons, 1997.
- [5] M. Felson, and E. Poulsen, "Simple indicators of crime by time of day", *International Journal of Forecasting*, Volume 19, Issue 4, p 595-601, 2003.
- [6] A. Graycar, "Local Government and Crime Prevention", *Proceeding of the character, Impact and Prevention of Crime in Regional Australia Conference*, Townsville, 2001.
- [7] T. Hagerstrand, "The domain of human geography", in Chorley, R. J. (ed.), *Directions in geography*. London: Methuen, pp. 67-87, 1973.

² 雖然犯罪時間關聯模式之檢出率都非常高，例如住宅竊盜案在關聯門檻0.6時，檢出率高達0.98，係因本研案將時間劃分為4個區域，且相鄰區域之權重亦給予0.5，次相鄰區域給予權重0.25，因此，每個時段大多會有權重，除非某個連續犯模型集中於某個特定時段，則次相鄰的時段才會出現權重為0，也因此關聯門檻為0.6及檢出率為0.98時，它的準確率只有0.06。

- [8] J. Han, and M. Kamber, *Data Mining: Concepts and Techniques*, Second Edition, Morgan Kaufmann, 2006.
- [9] H. C. Lee, *Crime Scene Investigation*, *Central Police University Press*, Taoyuan, Taiwan, ROC, 1994.
- [10] M. Palmiotto, "Crime Pattern Analysis: An Investigative Tool", in *Critical Issues in Criminal Investigation*, 2nd ed., Pilgrimage, 1988.
- [11] Y.-T. Peng, and J.-H. Wang, "Link Analysis Based on Webpage Co-occurrence Mining – a Case Study on a Notorious Gang Leader in Taiwan", *Proceedings of the IEEE International Conference on Intelligence and Security Informatics*, 2008.
- [12] G. Salton, *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*, Addison Wesley, 1989.
- [13] C. E. Shannon, and W. Weaver, *The Mathematical Theory of Communication*, *University of Illinois Press*, Urbana, IL, 1949.
- [14] B. E. Turvey, *Criminal Profiling, Third Edition: An Introduction to Behavioral Evidence Analysis*, *Academic Press*; 3 edition (April 30, 2008)
- [15] J.-H. Wang, B.-T. Lin, C.-C. Lin, "Application of the Vector Space Model on Criminal Record Retrieval", 1997 *國際科學警察學術研討會論文集*, 中央警察大學, 1997。
- [16] S. Wasserman, and K. Faust, *Social Network Analysis and Applications*, New York and Cambridge, 1994.
- [17] [http://en.wikipedia.org/wiki/Entropy_\(information_theory\)](http://en.wikipedia.org/wiki/Entropy_(information_theory)).
- [18] http://en.wikipedia.org/wiki/Locard's_exchange_principle.
- [19] http://en.wikipedia.org/wiki/Modus_operandi.
- [20] http://en.wikipedia.org/wiki/Social_network.

