

國立交通大學

管理學院（資訊管理學程）碩士班

碩 士 論 文

機場偷渡犯罪分析之資料探勘應用

Data Mining Application for Airport Illegal
Immigration Analysis

研 究 生：李權龍

指 導 教 授：劉敦仁 博士

中 華 民 國 九 十 七 年 六 月

機場偷渡犯罪分析之資料探勘應用

Data Mining Application for Airport Illegal
Immigration Analysis

研究生：李權龍

Student : Chuan-Lung Li

指導教授：劉敦仁 博士

Advisor : Dr. Duen-Ren Liu



A Thesis

Submitted to Institute of Information Management
College of Management

National Chiao Tung University

In Partial Fulfillment of the Requirements

For the Degree of

Master of Science

in

Information Management

June 2008

Hsinchu, Taiwan, the Republic of China

中華民國九十七年六月

中文摘要

機場偷渡犯罪分析之資料探勘技術應用

研究生：李權龍

指導教授：劉敦仁 博士

國立交通大學管理學院（資訊管理學程）碩士班

摘要

隨著交通工具便利性的提昇，國與國之間的距離逐漸縮短，世界各國之間的經濟與文化等交流活動日益頻繁，也使得大量合法或非法的移民人潮湧現，增加了各國對國境安全及機場治安上的顧慮。在以往的偷渡事件，多由「陸路」或「海路」進行，但隨著時代變遷，新穎的「空路」手法，除了提供安全、迅速的交通，並擁有以合法掩護非法犯罪之優勢，成功率大幅增高，使得偷渡行為更加猖獗囂張；且自美國九一一恐怖事件發生之後，各國莫不重視機場航空安全議題，綜上所述，本論文即以「空路」（機場偷渡）之問題作為探討重點。

本論文從相關文獻探討機場偷渡的行為，並分析人蛇集團常見的手法，進而導入線上即時分析處理（OLAP）架構及資料探勘（Data mining）技術的方法，針對機場偷渡行為作一分析處理。本研究以桃園機場（中正國際機場）查獲之偷渡案件為研究對象，建構OLAP分析架構，作為系統分析之主軸，並依據查緝人員實務上之需求，進行資料探勘分析之系統實作。本論文將目前的機場偷渡現象，以資料探勘技術進行分析、解讀偷渡行為之趨勢，以協助移民單位查緝人員得以從中獲得有用之資訊，並得以從勤務的規劃面及執行面上提供有效之參考與運用。

關鍵字：機場偷渡、資料倉儲、線上即時分析處理、資料探勘

英文摘要

Data Mining Application for Airport Illegal Immigration Analysis

Student : Chuan-Lung Li

Advisor : Dr. Duen-Ren Liu

Institute of Information Management

National Chiao Tung University

Abstract

As the convenience of transportation improving, the distance between countries is getting shorter. The economic and cultural communications between countries are frequently occurred that causes a majority of legal or illegal immigrants. This phenomenon raises the public security concern. In the past, most of the stowaways have succeeded by ways of ships or land transportations. However, the latest way to stow away is to board an aircraft with counterfeit or forge travel documents. The new method not only can provide safe and rapid advantages but also increase the rate of success. Therefore, it leads to more and more people trafficking. Since the September 11, 2001 attacks, the security of international airport becomes the most important issues for each country. Thus, the research objective of this dissertation mainly focuses on investigating the issues of using counterfeit or forge travel documents to board an aircraft in order to stow away.

In this research, literature survey is conducted to find the behaviors and methods of human trafficking. Moreover, this research employs OLAP and Data Mining to analyze the behaviors of illegal emigrants. This dissertation uses the illegal emigrants of the Taoyuan international airport (C.K.S. international airport) as the data set to conduct the analysis. The research first constructs the OLAP framework for illegal immigration analysis. Then, according to the immigration officers' practical demands, this research uses data mining to find the appearance of Airport illegal immigrations and understand the trend. Finally, this research concludes by suggesting a number of key issues that need to be addressed in the clampdown of human trafficking.

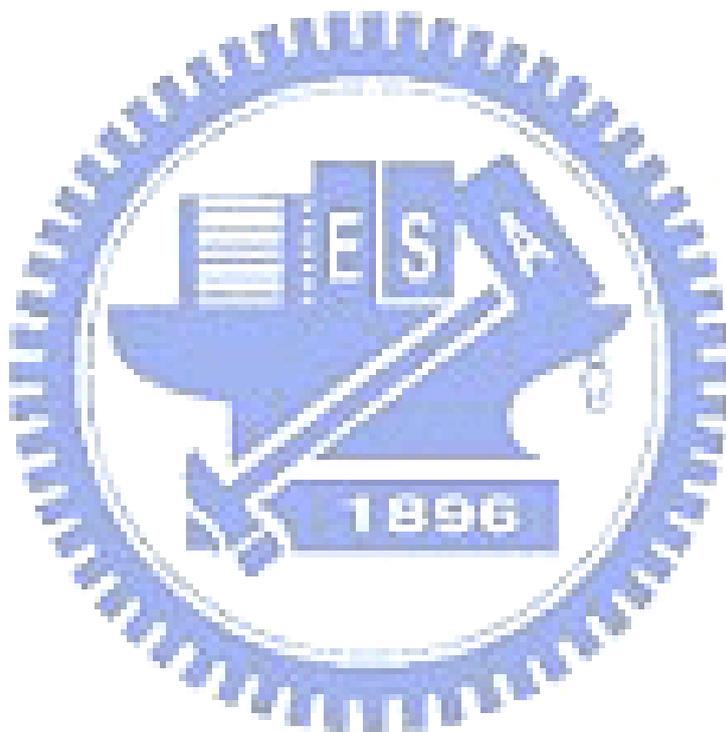
Keywords: Airport Illegal Immigration, Data Warehouse, OLAP, Data Mining

誌 謝

本論文能夠順利地完成，除了兩年來的辛苦以及努力之外，最重要的是要感謝周遭的師長、親友的指導與協助。在此，謹以最誠摯的謝意，感謝您們！

首先感謝我的指導教授劉敦仁教授，由於他熱忱的教學態度與豐富的知識經驗，使我能確立出研究的方向。再來，要感謝口試委員羅濟群教授、吳美玉博士對於論文內容之指正與建議，以及研究室同學、同事在論文寫作、口試期間時所給予的協助。

最後，最要感謝的是我的家人，在這段求學期間中對我的支持與關懷，讓我可以順利完成學業，感謝！



目 錄

中文摘要	I
英文摘要	II
誌 謝	III
目 錄	IV
表 目 錄	VII
圖 目 錄	VIII
一、緒論	1
1.1 研究背景及動機	1
1.2 研究目的	3
1.3 論文架構	5
二、文獻探討	7
2.1 偷渡	7
2.1.1 偷渡之定義	7
2.1.2 偷渡之目的	7
2.1.3 偷渡之交通方式	9
2.2 人蛇集團偷渡手法	10
2.2.1 人蛇集團	10
2.2.2 「空路」偷渡手法	11
2.3 資料倉儲 (DATA WAREHOUSE)	14
2.3.1 資料倉儲的定義	14
2.3.2 資料倉儲的用途	15
2.4 線上分析處理 (OLAP)	15
2.4.1 線上分析處理的定義	15
2.4.2 OLAP 與資料採礦的關係	16
2.5 資料探勘 (DATA MINING)	16
2.5.1 資料探勘之定義	16
2.5.2 資料探勘之五大功能	16
2.5.3 資料探勘之處理流程	17
2.5.4 關聯規則 (Association Rule)	18
2.5.4.1 關聯規則之定義	18
2.5.4.2 關聯規則之方法	19

2.5.4.3	關聯規則之應用領域：.....	20
2.5.5	分群 (Clustering)	20
2.5.5.1	分群之定義.....	20
2.5.5.2	分群之方法.....	21
2.5.5.3	分群之應用領域：.....	21
2.5.6	時間序列 (Time Series)	22
2.5.6.1	時間序列之定義.....	22
2.5.6.2	時間序列之方法.....	23
2.5.6.3	時間序列之應用領域：.....	23
三、	系統分析與設計.....	25
3.1	研究範圍.....	25
3.2	需求分析.....	25
3.3	系統架構.....	26
3.4	資料倉儲資料表設計.....	27
3.4.1	資料蒐集.....	27
3.4.2	資料預處理.....	28
3.4.3	資料探勘輸入變數型態及採礦模型角色處理.....	31
3.5	軟、硬體設備需求.....	32
四、	系統實作與展示.....	34
4.1	建立機場偷渡及班機起降維度資料表.....	34
4.2	CUBE 模式分析.....	34
4.3	樞紐分析表及圖表建立.....	35
4.3.1	國籍與偷渡目的地之樞紐分析.....	36
4.4	探勘主題.....	38
4.4.1	從偷渡方法、目的地，探勘特定國籍偷渡者之喜好度 (使用關聯規則).....	39
4.4.2	以關聯規則分析偷渡喜好 (使用關聯規則).....	43
4.4.2.1	「航班號碼」關聯規則分析.....	45
4.4.2.2	「航空代理公司」關聯規則分析.....	47
4.4.2.3	「機場」關聯規則分析.....	49
4.4.2.4	「國家」關聯規則分析.....	51
4.4.2.5	「假照國籍」關聯規則分析.....	53
4.4.3	偷渡客犯罪手法之趨勢 (使用群集).....	56
4.4.4	偷渡客選擇目的地機場的趨勢 (使用群集).....	61
4.4.5	偷渡客人數、時段之影響預測 (使用時間序列分析).....	64
4.4.5.1	偷渡客人數、時段之趨勢預測.....	64
4.4.5.2	經濟數據與失業率對於偷渡人數之影響預測.....	68
五、	結論與未來研究方向.....	72

5.1	結論.....	72
5.2	未來研究方向.....	74
參考文獻		76

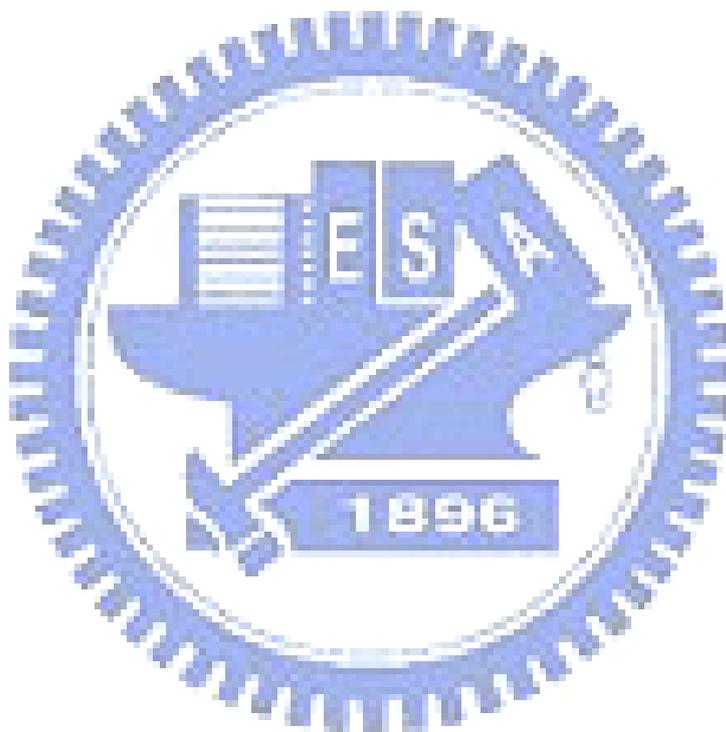


表 目 錄

表 1 民國 96 年 1-12 月份台灣桃園國際機場國際航線班機載客情形表	2
表 2 「輸入變數」適用之資料型別	32
表 3 關聯規則參數即輸入變數設定表	39
表 4 各國籍偷渡客人數分佈表（依照洲別分）	40
表 5 以關聯規則分析偷渡喜好之理由	43
表 6 以購物籃分析為主題之關聯規則參數設定	45
表 7 「航班號碼」關聯規則整理表	47
表 8 「航空代理公司」關聯規則整理表	48
表 9 「機場」關聯規則整理表	50
表 10 「國家」關聯規則整理表	53
表 11 「假照國籍」關聯規則整理表	56
表 12 偷渡客犯罪手法之趨勢（分群參數設定）	58
表 13 偷渡客選擇目的地的趨勢（分群參數設定）	62
表 14 偷渡客人數、時段之時間序列分析參數設定	65
表 15 經濟數據與失業率對於偷渡人數之影響預測參數設定	69



圖目錄

圖 1 台灣桃園機場國際班機航線圖	2
圖 2 AR 與 ART 方法比較	24
圖 3 系統分析架構圖	26
圖 4 資料庫架構圖	27
圖 5 樞紐分析圖	35
圖 6 樞紐分析表	36
圖 7 採國籍—偷渡目的地，採用「性別」偷渡樞紐分析表	37
圖 8 採國籍—偷渡目的地，採用「轉機」偷渡樞紐分析表	38
圖 9 關聯規則：支持度與項目集表	41
圖 10 關聯規則：「規則表」	41
圖 11 「航班號碼」關聯規則分析結果「相依性網路」圖	46
圖 12 「航空代理公司」關聯規則分析結果「相依性網路」圖	48
圖 13 「機場」關聯規則分析結果「相依性網路」圖	50
圖 14 「國家」關聯規則分析結果「相依性網路」圖	51
圖 15 「假照國籍」關聯規則分析結果「相依性網路」圖	54
圖 16 「假照國籍」關聯規則分析結果「相依性網路」右側放大圖	54
圖 17 「假照國籍」關聯規則分析結果「相依性網路」左側放大圖	55
圖 18 偷渡犯罪手法群集表，各群間連結強度（強度愈強，數字愈小）	59
圖 19 偷渡犯罪手法趨勢「群集設定檔」	60
圖 20 偷渡犯罪手法「群集辨識」	61
圖 21 偷渡客選擇目的地的趨勢「群集辨識」	62
圖 22 偷渡客選擇目的地的趨勢「群集設定檔」	64
圖 23 93/1/1-94/12/31 時間序列分析-1「絕對人數量」	66
圖 24 93/1/1-94/12/31 時間序列分析-2「相對人數量」	66
圖 25 93/1/1-94/12/3 每週偷渡人數分析	67
圖 26 美、加、日三國 93/1/1-94/12/31 偷渡時段分析	68
圖 27 美、加、日三國 93/1/1-94/12/3 偷渡人數與經濟數據之影響分析	70
圖 28 美、加、日三國 93/1/1-94/12/3 偷渡人數與失業率數據之影響分析 ...	71

一、緒論

1.1 研究背景及動機

台灣地區地理位置特殊，位居亞太交通要衝，戰略軍事及經濟地位重要性不言而喻。台灣除四面環海航運發達外，亦是國際重要空運航線交會處，由世界各地主要城市飛往北美、東北亞與東南亞的航線在此匯集，為亞太地區交通營運之樞紐。由於具備上述交通優勢，經濟發展快速，海峽對岸經由「海路」（海上偷渡）偷渡上岸來台打工之案件也屢見不鮮、防不勝防，且近十年來隨著開放外籍勞工政策與國際間交流日益頻繁，也使得台灣地區成為各國偷渡客淘金之天堂與轉飛他國之跳板。

近幾年來，世界各國的偷渡犯罪活動已經由以往常見之「海路」轉變為「空路」（機場偷渡）方式進行，「空路」偷渡方式，相對於「海路」提供了「長距離」、「安全舒適」的特性，且此種以合法交通工具掩護非法犯行的手法，已經讓各國國境安全受到極大之困擾及威脅。尤其自九一一恐怖攻擊事件以來，各國無不重視機場安全之議題，對於出入境之旅客莫不採取嚴格查驗之手段，以杜絕恐怖份子入侵國土攻擊。而機場偷渡案件防杜不力，不僅危及到國境安全，亦影響到國家於國際社會間之整體形象。

台灣桃園國際機場（前身為桃園中正國際機場），位於桃園縣大園鄉，主體共有 2 座航站大廈、28 條機場滑行道，為各國旅客來台首要進出之最重要的空運門戶。截止 95 年底止，飛航國際航線之國內航空公司共有 5 家，國外航空公司共計 38 家。航線橫跨亞、歐、美、澳、紐等區域，23 個國家[1]，航線圖如圖 1 所示。台灣桃園國際機場每日入出境旅客均數以萬計，2007 年載客情形表如表 1 所示。

資料探勘技術，已被企業採用且應用在專業領域上行之有年，不僅協助企業有效解決生存、轉型上之問題，更為企業創造豐碩之利潤。而機場執法機關，例如：保安部門及移民查驗部門，若能藉由資料探勘技術，挖掘出資料所隱藏的資訊，不僅可以協助其於勤務

制訂上之規畫面及執行面上之技巧改善，更能順應犯罪之潮流趨勢，得以掌握因應犯罪行為，與時俱進增強查緝作為。



圖1 台灣桃園機場國際班機航線圖
資料來源：[1]

表1 民國 96 年 1-12 月份台灣桃園國際機場國際航線班機載客情形表

中華民國 96 年台灣桃園國際機場國際航線班機載客情形表						
月別	入境		出境		合計	
	飛航班次	載客人數 (人)	飛航班次	載客人數 (人)	飛航班次	載客人數 (人)
總計	62,020	12,602,067	61,979	12,615,901	123,999	25,217,968
1 月	5,080	911,565	5,101	989,336	10,181	1,900,901
2 月	4,952	1,032,176	4,929	982,007	9,881	2,014,183

3月	5,169	1,025,385	5,197	1,068,117	10,366	2,093,502
4月	5,179	1,086,238	5,186	1,066,831	10,365	2,153,069
5月	5,272	1,019,764	5,274	1,049,365	10,546	2,069,129
6月	5,163	1,086,244	5,168	1,102,585	10,331	2,188,829
7月	5,451	1,154,027	5,433	1,163,953	10,884	2,317,980
8月	5,318	1,149,344	5,295	1,140,128	10,613	2,289,472
9月	5,141	1,061,000	5,100	1,018,045	10,241	2,079,045
10月	5,140	1,008,878	5,144	1,044,221	10,284	2,053,099
11月	4,932	1,012,484	4,933	981,609	9,865	1,994,093
12月	5,223	1,054,962	5,219	1,009,704	10,442	2,064,666

資料來源：[2]

1.2 研究目的

台灣地區由於地理位置特殊且交通、經濟發達，成為對岸偷渡者夢寐以求之目的地已是不爭之事實，本論文將研究之目的聚焦於「空路」之偷渡犯罪行為探勘，而採行「空路」相較於「海路」之偷渡犯罪有一極大不同點，即為「距離」不受地域限制，且以「合法」掩護「非法」之國際航空器實施犯行，故偷渡對象已不侷限為常見的大陸人民偷渡行為，而是屬於國際型態之偷渡行為。企圖來台入出境或轉機之外國偷渡客，則以「空路」直達台灣，或「空路」搭配「海路」或「陸路」迂迴抵達台灣或由台灣轉機到第三地，以遂行其偷渡目的。

於機場所查獲的眾多偷渡案件中，有絕大部分是查緝人員於為機場管制區中的轉機區內查獲，部份則是旅客於機場入出境查驗櫃檯辦理通關手續時查獲。而查緝單位若能利用資料探勘技術，整合相關數據資料，對有興趣之主題進行假設、探勘、驗證，則可協助

查緝單位作以下規劃：

1、對勤務規劃面而言：從已查獲之案件數據，利用資料探勘可將查緝人力資源分配在最需要之勤務上，例如：可針對特定期間，以時間序列找出偷渡熱門月份、日期、時段，加強編排特定航班查緝人力。或發掘出犯罪手法之趨勢改變，進而規劃於不同地點查緝偷渡，或編排相關之教育訓練，例如：某語系旅客偷渡人數劇增，應編排訓練查緝人員瞭解該語系旅客之語言、國家文化背景、護照辨識技巧等，以遏止偷渡問題繼續惡化。

2、對勤務執行面而言：查緝機場之偷渡犯罪案件時，旅客所出示之證件、登機證或其表情、服裝打扮等跡象有所異常，始會引起查驗人員進一步盤問。但轉機旅客因不需經過查驗櫃檯逐一查驗，故能降低被查獲的風險，此時若能透過資料探勘技術，分析出特定規則或趨勢時，例如：前往美國之旅客，若其由香港出發且未持台灣所核發之入出境許可證件，可懷疑其為偷渡客之關聯規則時，可採重點查驗，節省繁複過程之手續，快速過濾有問題之旅客。又若某一常偷渡美國之群集旅客，都為某特定國籍人士且搭乘某特定公司航號班機時，可從此特性解讀出是否有某種偷渡趨勢或航空公司職員是否已被人蛇集團收買之判讀等。

而從偷渡資料中找出其所隱藏、客觀之有用資訊，提出協助上述勤務規劃、執行上之分析建議，為本論文之目的。

1.3 論文架構

本論文共分成五章，編排之結構如下：

1、第一章「緒論」：

說明本論文研究的背景、動機、研究目的及架構。由於目前國內在防範機場偷渡行為研究的領域上，尚無一專文研討以資料探勘技術協助進行分析。而為瞭解機場偷渡犯罪行為現行之發展及未來之趨勢，並提供查緝單位於勤務規劃面及執行面之輔助參考，故以此切入作為研究主題做為此篇論文的動機。

2、第二章「文獻探討」：

主要介紹偷渡之定義、現狀及目前機場偷渡之犯罪手法；另外並介紹在進行資料分析時，所需採用的技術架構：資料倉儲(Data Warehouse)、線上分析處理(OLAP)與資料探勘(Data Mining)方法、分析工具等技術的相關文獻，以建置出本研究之目標。

3、第三章「系統分析與設計」：

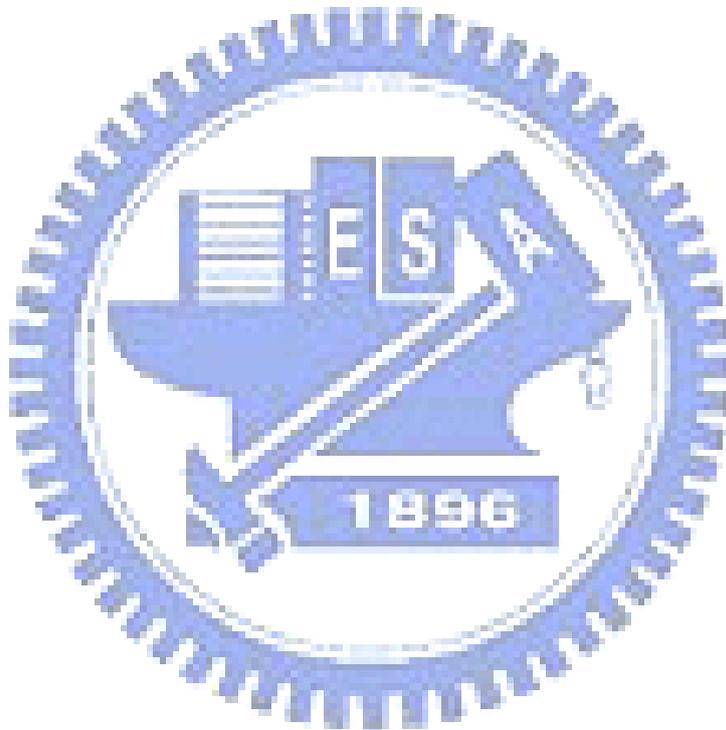
本章為實際以台灣桃園機場偷渡案件資料為研究標的，並說明出研究之範圍、需求分析、系統架構雛形及倉儲架構、其中亦包括資料的轉換與資料預處理問題及所需的軟、硬體設備需求。

4、第四章「系統實作與展示」：

將欲瞭解之主題透過資料探勘技術實作，做一內容之呈現與分析。以前一章建立的資料倉儲(Data Warehouse)及線上即時分析(OLAP)架構為基礎，搭配 SQL Server 2005 Analysis Service 工具，依據所欲分析的主題，採用多種不同的資料探勘(Data Mining)技術，並將結果以圖表呈現、解讀資料所隱藏的資訊。

5、第五章「結論與未來研究方向」:

希望能將本論文的研究成果，提供查緝單位以不同的面向瞭解目前偷渡行為所隱藏的資訊，以提供於勤務規劃面及執行面設計之輔助參考，並對於未來偷渡行為趨勢更能掌握。另一方面，也針對未來可進一步分析之資料層面方向作一闡述。



二、文獻探討

2.1 偷渡

2.1.1 偷渡之定義

「偷渡」一詞，依一九五七年「布魯塞爾偷渡公約」，「偷渡者」之定義為：「未得船東或船長（或其他主管船舶之人）之允許，潛入船上並隨船離開登船地點或港口之人」。多數組織均曾對「偷渡」之意涵進行詮釋，而大多數國家移民法之規定「凡未持有合法入出境證件，且未經主管機關許可，而企圖入出境者」，皆可稱之為「偷渡」；同時，各國移民法中對「偷渡者」之界定，亦非僅限於以船舶為之者，始可稱為「偷渡」，如持偽變造證件企圖入出境（由機場、港口、邊境關卡入出），亦可稱為「偷渡」[3]。

2.1.2 偷渡之目的

國際間所發生之偷渡行為，其之所以形成均有其客觀因素存在，不論其國籍為何？所採行之交通工具為何？可能發動之原因歸納如下：

- 1、經濟或教育因素：因國內經濟情勢不佳，多為求改善生活品質，且意圖長期定居國外工作，以賺取較為優渥之薪資為主。或為改善下一代教育、生活因素，而選擇偷渡等。改善經濟狀況為最常見之偷渡成因，此類非法移民多屬較為貧窮或文化素質低落之國家為主。
- 2、因違犯刑案或欠稅案件而潛逃國外：此類原因多為於國內犯下重大刑案或經濟案件、或為遭通緝之嫌疑犯、因逃漏稅等禁止出境之人員，因無法經由正常管道出入國境，故採取偷渡方式，以規避司法審判。

- 3、進行非法走私活動：此類為經常性短期往返於特定之兩地間，販運違法物品或農產品，以牟取暴利。而常見的重大違法販運物品為槍械或販毒。
- 4、遭受政治、文化、宗教等迫害：此類多為提倡民主運動人士、宗教團體聚眾運動或提倡主權獨立之活動，因政治情勢而遭受政府鎮壓等。例如：天安門、法輪功、西藏獨立事件等。
- 5、恐怖份子活動：跨國地下組織從事地下破壞行動，伺機破壞，為掩飾其身份以長期在目的國活動。
- 6、情報人員活動：為某國為蒐集特定國防或跨國商業間諜蒐集相關情資，必須掩飾身份以達成任務者所採用。
- 7、雇用殺手入出國境行兇：委託人為避免遭刑事司法單位察覺其犯行，遙控或雇用第三地經過特殊訓練之殺手或犯罪集團成員，以偷渡方式掩飾其身份，而有利其執行任務，而委託人得以製造與犯罪無關連之證明。
- 8、遭受拐騙或人口販運集團控制：人口販運集團為圖謀暴利，設局以精巧之謊言欺瞞特定身份之被害人，並以集團分工方式運作販運至國外。此類移民多以年紀較輕生活歷練單純、教育程度較差之對象居多，部分受害人則依其性別及年齡受到特定集團之控制利用，例如：重大災害地點之孤兒遭販嬰集團、年輕女性被利誘或脅迫賣淫等。
- 9、其他：因天災造成飢荒或人禍，諸如：屠殺、戰爭動亂原因而形成難民之身分，其被迫必須逃亡第三地以尋求庇護。

2.1.3 偷渡之交通方式

國際間進行之偷渡活動不外透過下列三種途徑交叉運用：一是經由「陸路」，即由陸鄰邊界越境偷渡入境；二是經由「海路」，其中有搭船由港口矇混入出境，亦有偷渡專船於海岸搶灘登陸；第三種是經由「空路」，亦即搭飛機偷渡，此模式亦有以持用假證照矇混闖關或以轉機方式偷渡第三國。因搭乘國際線航班有一定資格條件限制，故以無證照偷渡情況較少，反而是有偷渡客在入境偷渡目的國之前將身上所持有之證件(包括假證照)進行銷毀丟棄，然後以難民身分要求庇護，遂行偷渡之目的。

就我國現況而言，四面環海並無與他國有陸地交界，故無由「陸路」入出境偷渡之問題。而海岸線長、島嶼較多，海運交通便捷，為各鄰近各國「海路」偷渡之最佳目的地。「空路」是我國對外交通往來相當重要之途徑，由於「空路」較「海路」具快速便捷、長程交通之優點，相對之下，也漸漸成為我國防杜偷渡之工作重點 [4]。

偷渡犯罪已經從眾多論文、研究數據資料顯示，呈現有所謂的「專業化」、「組織化」、「國際化」之現象。從已查獲的案件中，不論其採用哪一種交通方式，都可以瞭解集團在整個偷渡過程中，事前皆經過詳細之規劃、跨國且階段性之專業化分工。多數偷渡犯罪之交通皆非採行直達目的地之方式為之，因安排偷渡之集團會視目的地國家偷渡之難易度及特定路線，搭配一種或多種的交通方式，有時還必須採行採迂迴之方式完成目的。其中「空路」是集團規劃長程地區最主要的交通方式，而「陸路」或「海路」則是與「空路」交叉搭配，先行於在目的地周遭國家登陸，再伺機而動以完成偷渡行為。「空路」偷渡行為相較於由「海路」、「陸路」偷渡行為之特性較為不同有如下幾點：

- 1、 「空路」偷渡除以劫機、未經申請降落來台外，多數均屬於以「合法交通」工具，行「非法偷渡」之行為。且考量之重點在於以「長距離」之目的地為主。在偷渡過程中也較具「安全」、「舒適」。中途發生交通事故意外的機率也相對較低。

- 2、 「空路」偷渡行為，必須具備出發地國家許可之出入境證件、目的國簽證、航班時間安排、登機手續等。此種方式事前作業較為繁雜，且每次同時進行偷渡之人數也較少。
- 3、 「空路」偷渡行為所花費之成本較為高昂，且事前作業準備證件時間較長、需有人協助指導入境查驗教戰手則、與當地雇主接頭等。上述項目均需有組織細密的人蛇集團所經營。而其他方式就不需如此繁複。以對岸人民採「海路」偷渡台灣為例，因兩岸距離相近不遠，幾乎一搜簡單的舢舨或漁船，甚或游泳就可以進行偷渡登陸，偷渡的專業分工程度可以說是簡單多了。

2.2 人蛇集團偷渡手法

2.2.1 人蛇集團

人蛇集團係指居間從事媒介、專門安排大陸地區或潛逃至該區人民非法偷渡至大陸以外地區的一群人所組合而成之犯罪組織，是一種跨國性之犯罪組織。俗稱所謂之「蛇頭」即屬「人蛇集團」成員之一，目前「人蛇集團」分子，除大陸地區人民外，更有台灣地區人民、及其他各外國籍之人員參與，彼此間合作，以「組織化」、「企業化」及「國際化」方式經營偷渡。被安排之偷渡者主要以漁船或其他各種船舶載運人蛇上岸，依付費多寡再由上岸地之集團分子接送至各地或安排工作；或由人蛇集團以各種管道取得我國或他國之他人護照、簽證，以冒領、偽造或變造方式將人蛇本人相片貼於護照之上，再以安排接駁班機、交換登機證或由護航之集團分子引領陪同等方式矇混闖關出境或轉機至他國[4]。

有關人蛇集團的結構與分工歸納如下：

- 1、 大蛇頭：是人口走私活動背後的投資者，同時也是整個走私活動的監督者，但通常不被偷渡者所認識。多數訪談受訪者認為大蛇頭是擁有權力、財力及聲望且人

際網絡廣泛、政商關係良好的人，他們大多身居人口走私活動的幕後，平時不一定住在大陸，一般人很難接觸。

- 2、小蛇頭：通常居住在中國大陸，是大蛇頭與偷渡者（顧客）之間的中間人，主要負責發掘顧客、過濾顧客以及收取偷渡者支付的頭期款。小蛇頭為了爭取業績，賺較多的工資，就必須努力拉客，甚至使用誇大虛無的言詞來說服那些有偷渡念頭但意志未堅的大陸民眾。

2.2.2 「空路」偷渡手法

入出國時重要的就是證件，基本上證件齊全有效，理論上皆可入出國，但若為了規避法律制裁或逃避治安單位的緝捕等目的而有入出國之必要時，就會先以不合法方式取得並持用入出境證件入出國，因而衍生犯罪行為。以下是不法證件的種類[5]：

- 1、冒領護照：一般俗稱為「人頭護照」，又稱「真的“假護照”」，即用本人的照片，冒用他人證件的基本資料，藉由欺騙公務員的方式，以申請為合法使用者。
- 2、冒用護照：直接冒名頂替持用他人的護照闖關，護照本身並未做任何變造或塗改。
- 3、偽造護照：整本護照從無生有，完全為假的，此種手法需有較專業的人士製作，且品質良莠不齊。
- 4、變造護照：以真的護照而將某些重要部份進行變造，如更換相片、更改護照基本資料等，變造部分由人蛇集團針對特定國家護照特性及需求決定。
- 5、變造簽證：專供被管制入境或申請不到入境簽證時所用。與變造護照手法相同，變造簽證通常是由申請成功之護照上貼換下來給申請不到簽證的護照上。

- 6、偽造簽證：目的與變造簽證相同，偽造整個簽證的難度較高，也較容易被當地查驗官員識破，故較少人使用。
- 7、變造重入境證：重入境證為專供外籍勞工來台申請不到入境簽證時所用，此手法與變造護照相同。
- 8、偽造重入境證：目的與變造重入境證相同，本國重入境證之防偽辨識效果不佳，故為人蛇集團有機可趁。
- 9、偽（變）造、冒領（用）入出境證：入出境證為因應兩岸政治情勢下，所專供大陸或港澳人士來台時持憑入出境的文件。其犯罪模式皆與偽（變）造、冒領（用）護照相仿。
- 10、空白護照：政府機關遺失或失竊之未經加工過的護照，易遭不法份子加工使用。
- 11、偽刻印查驗章戳：目的為以此偽造入出國紀錄，以取信查驗之移民官員。

由空路闖關模式計有以下幾種樣態[5]：

- 1、個人所為：自行闖關並無他人或藉由人蛇集團協助，亦未透過持有假證件之方式入出國。
- 2、親友協助掩護：由親朋好友一起同行，或持有效證照從旁協助翻譯掩護，或是持假證照假裝成家庭、夫妻等一起闖關，在查驗官員起疑時相互套招作偽證，以減少進一步被識破的危險。
- 3、組團掩護：利用旅行社團體旅遊之表象，掩飾其外表及言行，使查驗之移民官員降低戒心，進而矇混闖關。

- 4、人蛇集團從旁協助：由人蛇集團全程計畫偷渡闖關，若被查驗人員察覺可疑問話時，負責機場管制區之交通任務人員即會出面，以親友身分編理由或作偽證，以求達成順利通關的目的。
- 5、執法者縱容包庇：偷渡者或人蛇集團以賄賂或要脅等不法手段，要求執法人員對其集團所掩護之偷渡客於入出國境提供協助，參與其集團偷渡行為之分工。
- 6、迂迴偷渡：又稱「多角偷渡」，偷渡者原計劃偷渡至 A 國，但卻採迂迴第三地方式達成，雖然浪費時間和金錢，但卻可以選擇一個較容易偷渡成功之地點，減少入境國的懷疑，進而達到入境該國的目的。
- 7、過境轉機偷渡：此模式目的在藉助於經過所持用的假護照國，用以證明持照人剛由發照國合法出境，間接證明該護照是真的，以降低被入境國懷疑、查獲的機率。
- 8、合法掩護非法：以旅遊、商務考察、出國留學、結婚、依親、探病 等名義為主，再逾期停留不歸。
- 9、裝扮難民模式：這是大陸人士及東南亞各國想進入美、加等國最常用的招式。先造成已入境的事實，再以要求難民保護的方式交保候傳，等待司法程序。這時人蛇集團安排「捍衛人權」的人士保出，再棄保達到入境停留的目的。
- 10、外勞出境未登機回流：外勞進入出境查驗管制區後，以物品放在外頭忘記拿，或是要把物品交給在外頭的朋友等藉口要求退關，而在走出管制區後即不再回來。
- 11、直接闖關：查驗員發現入境旅客身分有疑慮，做進一步確認時，該旅客拔腿就跑。

2.3 資料倉儲 (Data Warehouse)

2.3.1 資料倉儲的定義

資料倉儲的概念首次出現於 1988 年 Barry Devlin 與 Paul Murphy 兩位學者的文章中，他們認為除了原有的交易系統之外，企業必須建置一個隔離的唯讀 (Read Only) 資料庫，整合所有作業資料與歷史資料，以提供給企業內部的專業決策人員之用。這個概念一直到 1991 年由 IBM 提出所謂的「資訊倉儲 (Information Warehouse)」後正式有商業化的產品推出。至於「資料倉儲」這個名詞則是由 Bill Inmon 於 1993 年提出 (所以資料倉儲誕生至今才十幾歲左右)，其為資料倉儲的資料庫設計特性進行定義，認為資料倉儲必須具備以下特性：主題導向 (Subject-Oriented)、整合性資訊 (Integrated)、能呈現時間變異 (Time-Variant)、非揮發性 (Non-Volatile) [6]。Inmon 首次點出了資料倉儲與一般交易系統設計邏輯的最大差異，同時也考量到「時間」對於資料倉儲檢視決策資訊時的重要性。Ralph Kimball 於資料倉儲中提出包括「星狀架構 (Star Schema)」、「雪花狀架構 (Snowflake Schema)」、維度資料架構、資料市集 (Data Mart) 等概念。

資料倉儲 (Data Warehouse) 廣義的說是指利用資訊科技所提供的大量資料倉儲及資料快速分析能力，使用邏輯分析、統計、彙總運算，將大量的營運交易資料建立成一個整合性、彙總性的資料庫系統，用來協助企業訂定正確及精準的營運決策。藉由整合企業的內部資料為基礎，加上結合各種不同的外部資料，透過邏輯運算的分析、彙總...等多面向處理，將資料轉換成有用的、策略性的資訊。提供企業決策者一個完整的、精確的方向，來支援決策的制訂，藉此提昇企業競爭力 [7]。

2.3.2 資料倉儲的用途

資料倉儲最重要的用途如下[7]：

- 1、提高企業競爭能力，降低成本、提高顧客滿意度，創造更高利潤：資料倉儲的主要用途在塑造企業一個經營決策分析的作業環境，藉此來改善企業整體的營運及強化其競爭力，以達到企業營運目的。
- 2、協助企業由正確的人在正確的時間來執行正確及有效的事：「資料倉儲」對於企業的最大貢獻在於「執行效率」，能適時地提供決策人員最需要的支援資訊。
- 3、作為企業高層的決策「智庫」：資料倉儲對於企業的功能運作，可說是一種循環進化的過程。使用資料倉儲的分析統計結果，可藉此擬定出企業的行銷業務策略，再將此行銷計畫實際付諸執行結果回饋到「資料倉儲」系統，執行更進一步的統計分析，使用此種方式讓行動／學習兩大作業不斷進化循環，企業的競爭力自然會與日俱增了。

2.4 線上分析處理 (OLAP)

2.4.1 線上分析處理的定義

線上分析處理 (On-Line Analytical Processing, OLAP) 是一種具有即時分析儲存在資料倉儲 (Data Warehouse) 中各類型資料的軟體分析及處理技術。OLAP 最重要的設計原則為透過「快速化」、「交談化」、「線上處理」的人機介面，運用相同資料庫的資料而能提供各種不同的呈現方式，供各個不同的應用層面來使用，使其具備透析資料反映出來資訊的能力[7]。OLAP 其技術核心，為採用「維度」(Dimension) 之概念，可供 OLAP 使用者由多維度的觀點來觀察並分析資料，且具備彙總迅速之優點，如此一來，便可使資料倉儲架構有效地運用於線上分析，進而對使用者查詢做快速地回應。

2.4.2 OLAP 與資料採礦的關係

所謂 OLAP 乃指由資料庫所連結出來的線上查詢分析程序。Data Mining 用在產生假設，OLAP 則用於查證假設。簡單來說，OLAP 是由使用者所主導，使用者先有一些假設，然後利用 OLAP 來查證假設是否成立；而 Data Mining 則是來幫助使用者產生假設。Data Mining 常能採礦出超越歸納範圍的關係，但 OLAP 僅能利用人工查詢及視覺化的報表來確認某些關係，是以 Data Mining 此種自動找出甚或不會被懷疑過的資料型樣與關係的特性，事實上已超越了我們經驗、教育、想像力的限制，OLAP 可以和 Data Mining 互補，但這項特性是 Data Mining 無法被 OLAP 所取代的[8]。

2.5 資料探勘 (Data Mining)

2.5.1 資料探勘之定義

資料探勘是近年來資料庫應用領域中相當熱門的議題。乃指在資料庫中，利用各種分析方法與技術，將過去所累積的大量歷史資料中，進行分析、歸納與整合等工作，以萃取出有用的資訊，找出有意義且令人有興趣的樣式 (Interesting Patterns)，提供企業管理階層在進行決策時之參考依據。資料探勘是指找尋隱藏在資料中的訊息，如「趨勢 (Trend)」、「特徵 (Pattern)」及「相關性 (Relation)」的過程，也就是從資料中發掘資訊或知識。

事實上，資料探勘並不只是一種技術或是一套軟體，而是一種結合數種專業技術的應用。資料探勘並非無所不能，只是從資料中發掘出各種假設 (Hypothesis)，也無法判斷這些假設對你的價值[8]。

2.5.2 資料探勘之五大功能

資料探勘功能一般而言包含五大項，內容多為以技術較為成熟之計量及統計分析為主之方法[8]。

- 1、分類 (Classification)：依照分析對象的屬性來進行分門別類，並加以定義。
- 2、推估 (Estimation)：依據既有連續性數值之相關屬性資料，來推估獲得某一屬性未知之值。
- 3、預測 (Prediction)：根據要進行分析之屬性其過去之觀察值，來推估該屬性未來之值。
- 4、關聯分組 (Affinity grouping)：從所有物件進行關聯分析，以決定物件間之關聯情形進而分組。
- 5、同質分組 (Clustering)：將母體中區隔出較具同質性之群組 (clusters)，目的是要將各組之間的差異辨識出來，而組內的樣本最相似，並對個別組內之相似樣本進行揀選。

2.5.3 資料探勘之處理流程

資料探勘的過程會隨著不同領域的應用而不同，針對不同問題需求所發展出來的資料探勘過程也會有差異化的存在，如資料的完整程度、專業人員支援的程度等。都會對建立資料探勘過程有所影響，也因此造成資料探勘在各不同領域間運用規劃整個流程上產生差異性，即使是同一產業，也會因為不同分析技術結合不同涉入程度的專業知識，而產生明顯的差異性，因此對於資料探勘過程的系統化、標準化就顯得格外重要，如此一來不僅可以較容易跨領域應用，也可以結合不同的專業知識，發揮資料探勘的真正精神[8]。

資料探勘完整的進行步驟，依序為「理解資料與進行的工作」→「獲取相關知識與技術 (Acquisition)」→「整合與查核資料 (Integration and Checking)」→「去除錯誤或不一致的資料 (Data Cleaning)」→「發展模式與假設 (Model and Hypothesis Development)」→「實際資料探勘工作」→「測試與檢核所探勘的資料 (Testing and Verification)」→「解

釋與使用資料 (Interpretation and Use)」。由上述步驟可看出，資料探勘牽涉了大量的準備工作與規劃過程，事實上許多專家皆認為整套資料探勘的進行有 80% 的時間精力是花費在資料前置作業階段，其中包含資料的淨化與格式轉換甚或表格的連結。由此可知資料探勘只是資料探勘過程中的一個步驟而已，在進行此步驟前還有許多的工作要先完成。

CRISP-DM 是 (Cross-Industry Standard Process for Data Mining) 的簡稱，此為歐洲委員會與幾家在資料探勘應用上有經驗的公司共同籌劃組織的一個特別小組，其提出 CRISP-DM 模型，乃以方法學的角度強調資料探勘的方法與步驟，其認為完整的資料探勘過程，不能只針對在資料整理、資料呈現、資料分析以及建構模式上，仍需要對企業的需求問題進行瞭解，以及後期對模式的平價與模式的延伸應用都是不可或缺的。其所提出之六個階段，依序簡介如下：「定義商業問題 (Business Understanding)」、「資料理解 (Data Understanding)」、「資料預處理 (Data Preparation)」、「建立模型 (Modeling)」、「評價和解釋 (Evaluation)」、「實施 (Deployment)」 [8]。

2.5.4 關聯規則 (Association Rule)

2.5.4.1 關聯規則之定義

關聯規則 (Association Rule) 是由 R.Agrawal 與 R.Srikant 於 1994 年提出的一種高效率資料探勘方法 Apriori 而來 [14]。乃指從交易資料庫中、挖掘及尋找具有特定關聯的規則，亦即找出資料與資料間的關聯性。

關聯規則其所運用的原理，即為「條件機率」。而欲判斷哪些規則是有意義且有效的，則應觀察下列兩個指標：

- 1、信心水準 (Confidence)：指此關聯規則的準確度有多少，若從條件機率的公式來看，可以解讀為某條件 A 發生下，B 條件也會發生的可能性。

- 2、支持度 (Support)：信心水準高雖然表示此規則具有高準確度，但如果此規則交易的次數相當少時，那麼要運用這條規則時，就有待商榷。

信心水準及支持度可以用來判斷規則的有效程度，也可以在演算法進行時縮減規則的數量。

2.5.4.2 關聯規則之方法

常見的關聯規則方法為 Apriori 演算法，其執行步驟如下[14]：

- 1、首先，需訂定最小支持度及最小信心水準。
- 2、Apriori 演算法因使用了候選物項集合的觀念，首先產生出物項集合，稱為候選物項集合，若候選物項集合的支持度大於或等於最小支持度，則該候選物項集合為高頻物項集合 (Large Itemset)。
- 3、在 Apriori 演算法的過程中，首先由資料庫讀入所有的交易，得出候選單物項集合 (Candidate 1-itemset) 的支持度，再找出高頻單物項集合 (Large 1-itemset)，並利用這些高頻單物項集合的結合，產生候選 2 物項集合 (Candidate 2-itemset)。
- 4、再掃描資料庫，得出候選 2 物項集合的支持度之後，再找出高頻 2 物項集合，並利用這些高頻 2 物項集合的結合，產生候選 3 物項集合。
- 5、重複掃描資料庫與最小支持度比較，產生高頻物項集合，再結合產生下一級候選物項集合，直到不再結合產生出新的候選物項集合為止。

2.5.4.3 關聯規則之應用領域：

關聯規則在資料探勘技術中，普遍應用在如下幾類[9]：

- 1、交叉銷售分析：對特定企業來說，客戶每次去購物時可能只會購買 1~2 種商品(像是金融、保險、電信等)，運用關聯規則，可推薦客戶可能喜好的商品。
- 2、網路瀏覽者行為分析：此種應用可提升網站的使用程度，分析的標的就是所有使用者瀏覽的連結資料，透過使用者的 Session，可進行所謂的購物籃方式分析。
- 3、繳款行為分析：繳款行為也是一種序列，不管是超商、信用卡繳款還是分期付款，我們都可以將歷史繳款記錄當作一系列的序列，從中可以從中找出客戶的繳款模式，找出其中的關聯性。

2.5.5 分群 (Clustering)

2.5.5.1 分群之定義

分群為衡量事物之間之「相似性」，為依據樣本在幾何空間上的「距離」來判斷。與樣本「相對距離」較近的，我們說它們的「相似程度」越高，可歸併為同一組。分群與分類非常的相似，同樣也是希望透過由樣本的分組，因而尋找其中的差異之處。其與「分類」所不同之處有兩點：首先，群集分析的分群方式並不需要預先指定一指標變數；再來，群集分析屬於一種無母數分析方法，所以並沒有非常嚴謹的數理依據，當然也不需假設母群體為常態分配。

在眾多的多變數分析方法中，群集分析法是比較簡單的一種，統計學家通常應用群集分析法來對資料作簡化的工作及分類，也就是把相似的個體(觀測物)歸類於一群。不過，究竟相似的標準為何？多相似才能歸為同一群，都有待探討才能運用在實際狀況上。

2.5.5.2 分群之方法

目前最常用的群集分析演算技術分成分割式、階層式群集演算法兩大類：

- 1、 分割式群集演算法：主要分析方法為找到分群中心最短距離的應用，主要分析點為找出群集大小相似的群集。常用的演算法為 K-平均法 (K-means)、期望值最大化 (Expectation maximization, EM)。
- 2、 階層式群集演算法：主要演算方式是將資料庫內的資料使用不同階層式的樹狀架構 (Tree Structure) 來分析及運算的階層式群集演算法。

本研究所採用之分群法，乃採「期望值最大化」EM 方法。EM 方法乃使用機率函數進行量測對象屬於哪一個群集，而非如 K-means 所採行嚴格的距離函數。EM 算法不是為每一個維度選擇一個點，然後計算距離，而是把每一個維度當作一個鐘形曲線，計算平均差及標準差。當一個點落到鐘型曲線內的時候，會以某一個機率分配給某一個群集。而因為每一個群集的曲線可以重疊，所以每一點可以屬於多個不同的群集，每一個群集有不同的機率。我們也稱這種技術為「柔性群集」[15]。此外，EM 方法採用 Z-score 做為正規化資料的方法，輸入的變數過多將導致計算過程造成系統很大的負擔[12]。

2.5.5.3 分群之應用領域：

分群在眾多的資料探勘技術中，依據其挖掘規則的型態為屬於「非監督式學習」的方式，而採用此方法的分析人員為企圖從資料中找出一些隱含的現象或探索資料內部結構關係，因為從資料中所呈現的「趨勢」，往往是決策者有興趣的部份，而其在實務的應用上，可應用在如下幾類[9]：

- 1、 市場區隔／顧客行為區隔：可從既有之客戶資料來找出客戶間自然的相似結構關係，並根據群集來設定市場區隔策略，最重要的是，群集演算法可以明確地標示

出每一個客戶是屬於哪一個群集，這對於行銷人員篩選名單或是設計客製化行銷時都相當有幫助。

- 2、 詐欺偵測：找出存在於顧客間的常態模式，相對來說，如果某客戶的行為偏離了這些常態模式，就代表為異常的極端案例。在保險業中，可以利用群集演算法找出客戶理賠的群集，偏離常態理賠模式的客戶，就可能是潛在的保險濫用者。
- 3、 製程問題排除：在製程上經過同樣處理程序以及製程條件的商品也會有相似的品質，在製成過程中會進行數十道不等的品管數值偵測，透過這些品管數值的分群，可以找出品管數值之間的相關性，以及瑕疵產品分佈的常態模式，以找出改良製程的關鍵因素。

2.5.6 時間序列 (Time Series)

2.5.6.1 時間序列之定義

經由不同時間反覆量測所得到的值或事件的序列組成。這些值通常是在一個相等時間間隔區分（例如：小時、日、週）所量測而得。時間序列常應用在股票市場分析、經濟及銷售預測、預算分析、效用研究、自然現象預測等[15]。

要使用時間序列來預測連續數值的規律性，就要將規律性分成兩個部分來看，一是趨勢 (Trend)、另一為季節性 (Seasonality)。「趨勢」是跟隨著時間變化而顯示數值增減的量，例如：股票分析常用的移動平均法 (Moving Average)。「季節性」則是根據某固定週期循環而得的規律，例如：大陸東北地區的一到三月份，降雨量是一年最高的。而建立時間序列模型的第一步應先找出「趨勢」，接著再找出「季節性」，然後將趨勢及季節性移除使時間序列呈現穩定的狀態，以建立預測的模型自我迴歸。

2.5.6.2 時間序列之方法

目前最常用的時間序列演算技術有下列兩種[9]：

- 1、自我迴歸 (AR: Auto Regression)：乃指時間序列中的每個值，都是距離現值前幾期值的線性組合，透過自我迴歸找出迴歸係數，以重新建立整個時間序列的方程式。數學式如下：

$$y_t = f_1 y_{t-1} + e_t$$

y_t ：時間 t 的觀測值

f_1 ：一階自我迴歸係數

y_{t-1} ：時間 $(t-1)$ 期，亦即前一期的觀測值

e_t ：誤差項與係數

- 2、自我迴歸樹 (Auto Regression Tree, ART)：乃指以決策樹為主體的自我迴歸，透過樹狀規則分枝節點，在分枝節點中各自產生相對應的 AR 方程式，如圖 2 所示可避免傳統 AR 僅以單一方程式描述時間序列的缺點，提高預測的準確性。而傳統決策樹是透過規則分枝，以找尋出資料亂度 (Entropy) 最小的變數，並產生分枝節點。而自我迴歸樹 (ART) 是利用決策樹的規則分枝找出能讓時間序列最為平穩的分枝點，然後在每個分支以 AR 方式產生相對應的方程式。這樣的作法比傳統的 AR 的結果在預測上準確許多[10]。

2.5.6.3 時間序列之應用領域：

時間序列分析主要的目的是用來預測未來的數值。以常見的製造、生產業為例，工廠的接單，除需滿足客戶訂單的出貨要求，還需有原料安全庫存的考量，故若能預測未來幾個月、甚至季、年的需求量，便可以掌握製造效率，維持庫存物料。尤其在現今原物料飛漲的年代，更需藉由時間序列分析來掌握未來的商機。而以下列舉常應用的領域：

- 1、股市、金融投資：例如：投資人重視金融商品價格隨時間變化的趨勢，經由時間序列分析，低買高賣，從中獲取相當利益。
- 2、商業銷售、業績分析：例如：零售業可經由歷史資料，透過時間序列分析，掌握安全庫存數量及分析顧客消費時段尖峰編排服務人員班表。
- 3、生產、製造業預測分析：例如：生產、製造業等工廠之 ERP、CRM 系統可利用時間序列分析，對生產、客戶之資料預測未來行為，以在商場洞燭機先。

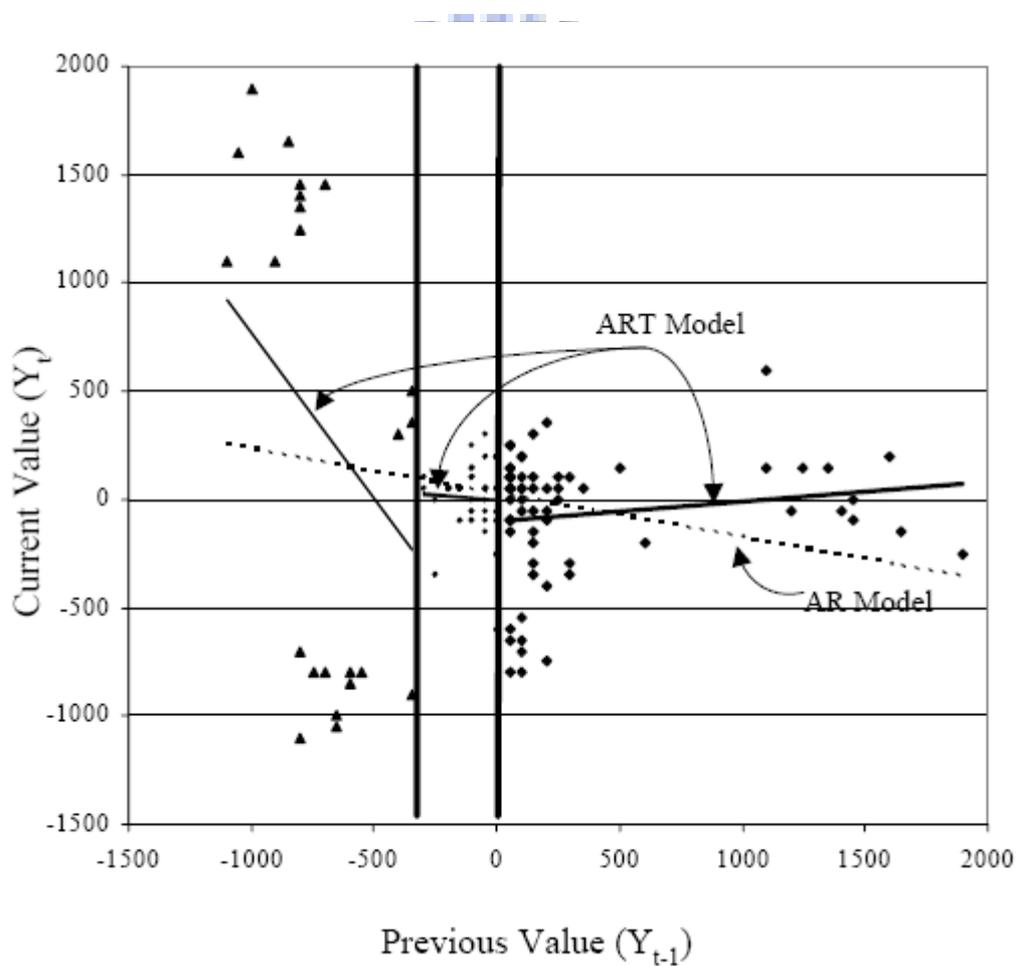


圖2 AR 與 ART 方法比較

資料來源：[13]

三、系統分析與設計

3.1 研究範圍

本研究係以台灣桃園機場（前身為桃園中正國際機場）遭緝獲之偷渡嫌疑犯為系統實作之研究對象，本論文研究方向偏重於偷渡客之偷渡行為與班機、目的地等資訊分析，以利查緝人員瞭解機場偷渡行為之趨勢，進而思考對策防杜並有效率分配查緝資源。

本研究實作資料蒐集範圍如下：

- 1、 時間：93/01/01~94/12/31。
- 2、 對象：遭查獲為人蛇集團成員或偷渡客。
- 3、 地點：機場管制區內（含第一、二航廈入出境查驗櫃臺、轉機室、登機室）。

3.2 需求分析

本論文將需求分析定位在機場查緝人員面對大量旅客入出境機場之狀態下，如何透過偷渡犯罪之行為分析，以因應於勤務編排的規劃層面與執行層面作為需求起點。故分析需求為先假設牽涉勤務之層面因素有哪些，進而對資料驗證以找出資料中隱藏之資訊。預定進行之探勘需求如下：

- 1、 從偷渡方法、目的地，探勘特定國籍偷渡者之喜好度（使用關聯規則）
- 2、 偷渡客對「航班號碼」、「航空公司」、「航空代理公司」、「機場」、「國家」、「假照國籍」等喜好分析情形（使用關聯規則）

3、偷渡客犯罪手法之趨勢發展研判（使用群集分析）

4、偷渡客人數、時段之影響預測（使用時間序列分析）

3.3 系統架構

本研究係將偷渡犯罪之資料從會議所附之紙本資料由資料庫設計架構後輸入，而為利未來資料庫之延展性考量，仍採取建立 Data Warehouse 架構，並配合 SQL Server 2005 Analysis Service 建置 OLAP 及使用 Data Mining 工具挖掘出本次研究所需資訊。如圖 3 所示，首先我們會使用關聯規則找出偷渡活動的規則，接著以分群方法找出偷渡活動的趨勢，最後，以時間序列分析偷渡的頻率，並將結果以相關圖表展示及專家意見，以提供勤務規劃人員決策參考。

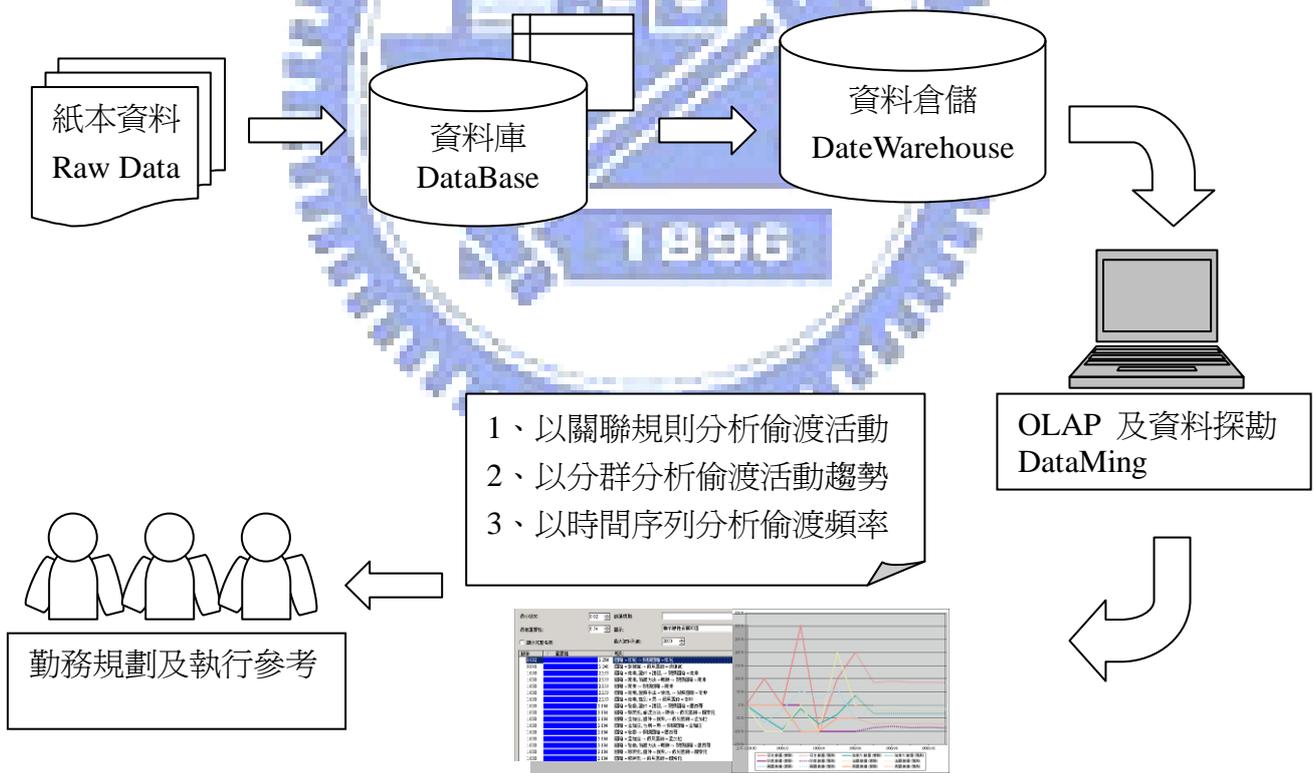


圖3 系統分析架構圖

資料來源：本研究整理

3.4 資料倉儲資料表設計

3.4.1 資料蒐集

因機場偷渡資料為以會議附件紙本文件方式記載，並非由一特定系統受理輸入，故進行分析前除需輸入電腦外，尚須進行資料錯誤修正及配合分析需求設計為特定資料之輸入格式。故此處如圖 4 所示我們先將偷渡資料依照內容及分析需求整理出一資料庫架構，資料庫架構內含 2 個主要資料表。

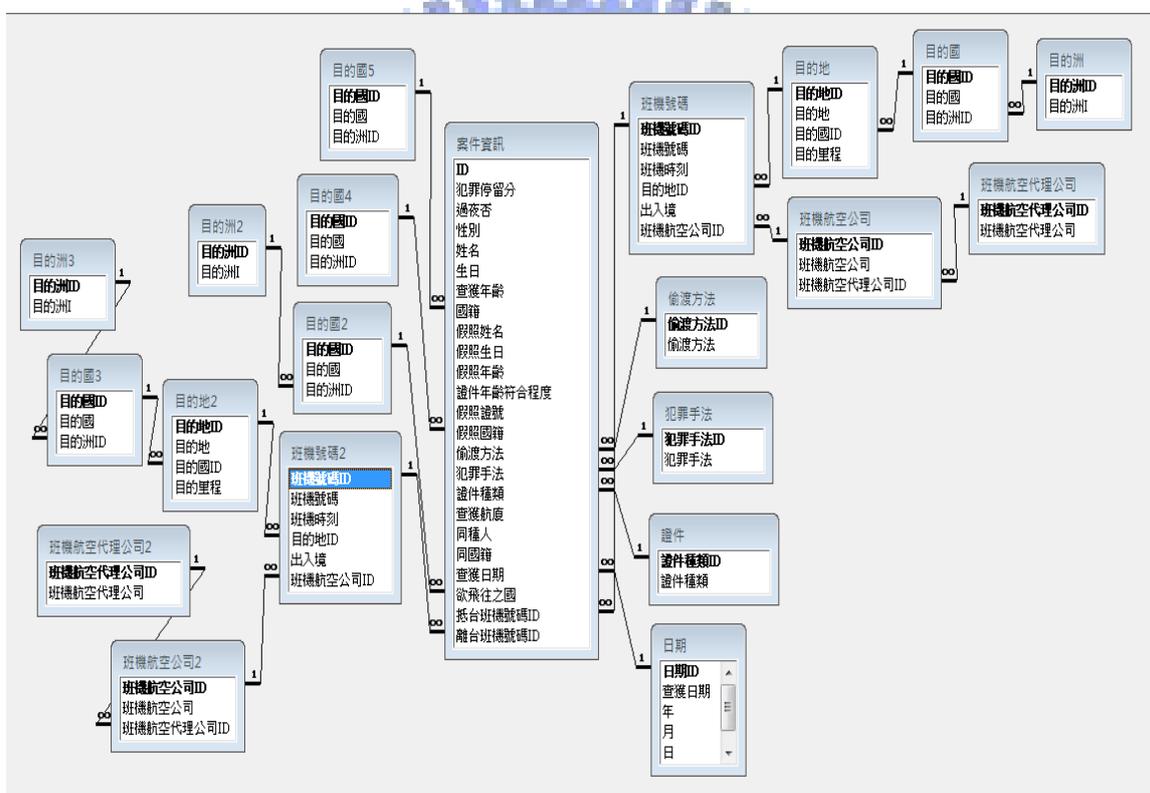


圖4 資料庫架構圖

資料來源：本研究整理

1. 案件資訊：記載查獲案件時之相關資訊，計有偷渡客「姓名」、「性別」、「生日」、「年齡」、「國籍」、「犯罪手法」；持用假證件資料之、「假照姓名」、「假照生日」、「假

照年齡」、「假照證號」、「欲偷渡之地點」等。

2. 班機資訊：記載偷渡客所搭乘之班機資訊，計有「班機號碼」、「出入境別」、「班機時刻」、「目的地」、「目的國」、「航空公司」、「航空代理公司」等。

3.4.2 資料預處理

在資料蒐集過程中，如前所述因原始之紙本資料偶有錯漏，必須經過人工重新核對資料之正確性及配合分析資料需求，對資料做統一性之修訂，另外還必須針對分析軟體架構所需，將資料進行轉換以符合資料探勘模型所需，例如：分群演算法中之類別變數處理、時序演算法所需之資料格式等。以下列舉所要進行之資料修正設計如下：

- 1、「年齡」：此定義指偷渡犯之年齡，以查獲日期為計算起始點。年齡等於「西元查獲日期」減去「西元出生日期」。
- 2、「年齡符合程度」：此定義為偷渡客實際之年齡與其所持假證件上所登錄之年齡差異值。該值等於「年齡」減去「假照年齡」，並取絕對值。
- 3、「偷渡犯罪手法」：此定義為偷渡客於機場管制區進行犯罪行為時，以何種偷渡方法策略、持用何種證件及其型態被查獲。作法是將原本為文字敘述部份，先歸類為「類別變數」，再依據犯罪手法之實際上之難易、進階程度，主觀轉換為有次序性的數字類別，以供資料探勘使用。我們將犯罪手法又細分為以下三類，分述如下：
 - (1)、偷渡方法：分為轉機（來台轉機至第三地）、出境（由台灣經由查驗出境）兩種。
 - (2)、持有證件：此部分依照被線上查緝人員查獲時，所持有之證件為準。並依所

持之證件與作用之重要性，訂出代表的數字類別，數字愈小，代表其重要性相對偏低，也易被辨識出為偷渡客。持有「許可證」，代表數字為 1；「綠卡」，代表數字為 2；「簽證」，代表數字為 3；「護照」，代表數字為 4；「護照 + 簽證」，代表數字為 5；「登機證」為 6；「登機證 + 查驗章」為 7。

- (3)、證件手法：此部分依照其所持證件之狀態，訂出代表的數字類別，數字愈大，代表其重要性相對偏高，也較不易被辨識出為偷渡客。證件若為「偽造」（證件整個為假），代表數字為 1；「變造」（亦即證件為真，但一部份為假造），代表數字為 2；「交付」（亦即協助偷渡之人蛇集團成員，其交付他人證件給予偷渡客之犯罪行為），代表數字為 3；「掩護」（亦即協助偷渡之人蛇集團成員，其因受盤查而承認其犯罪行為），代表數字為 4；「冒用」（亦即以他人證件矇混申辦證件成功，並持有而進入機場管制區），代表數字為 5。

- 4、班機「離開台灣時間」及「抵達台灣時間」：班機「離開台灣時間」乃指偷渡客由台灣出境或轉機之班機出發至他國機場之班機起飛時間；「抵達台灣時間」乃指偷渡客由他國機場出境來台之班機抵達時間。此部份之資料將略過班機因天候或人為因素延誤之起降時間，均以機場原訂公布之班機時刻表律定為準，資料格式的單位為 hh:mm。這樣的作法乃考量人蛇集團或偷渡客於安排偷渡時，並未能於其偷渡計畫中預知偷渡當日之天候、人為延誤因素。據此，故不考慮以實際班機抵達、離台之時間為分析之標的。另外，在資料的設計上我們考量偷渡行為有兩種形式，一是當日由國外機場來台轉機，另一是來台入境後由人蛇集團再安排出境。當日由國外機場來台轉機之偷渡客，可由其所持之機票、護照等出境資料，獲得其「抵達台灣時間」。但若被查獲之偷渡客屬於早於數日前即抵達台灣後始出境者，由於查緝人員無法確定其原始「抵達台灣時間」，故此處律定將其當初之「抵達台灣時間」設定為被查獲當日之「離開台灣時間」，以解決資料遺漏值之部分。

- 5、「犯罪停留時間」：此定義為從偷渡客搭乘班機抵達台灣時起算，至搭機離開台灣為止。單位為「分」。而偷渡客若係以過境轉機來台，停留時間即為「班機離開台灣時間」減去「抵達台灣時間」。若其從台灣出境查驗櫃臺通關者，我們統一將其預定的「班機離開台灣時間」往前推 2 個小時，亦即 120 分鐘做為其犯罪停留時間。此概念乃參考航空公司作業慣例，多於班機起飛前 2 個小時左右開放登機劃位，也較合乎實際狀況。
- 6、「出發地」及「目的地」：此定義以抵台或離台航班之「機場」所在「城市別」為地點，並非指偷渡客最終之目的地，因部分偷渡客即使採「空路」方式到航班目的地後，仍會採迂迴方式再以其他方式轉往真正的目的地，故非本論文研究探討之範圍。此處我們將「城市別」設計為「區域」的維度屬性。「區域」之維度內容，其下計有「洲別」、「國別」、「城市別」三種。若偷渡客搭機來台之「出發地」無法判斷，識別之依據首先為根據其班機航次出發位置查詢地點。若查無此資料，則依出境證件之查驗章紀錄，而最終仍無法辨識時，則視為由台灣出境，而非來台轉機之偷渡客。另外，遇航班之「出發地」到「目的地」中途有轉機情形時，則略過中途轉機地；亦即中途有停留到台灣以外的第三地載客才飛往最終的目的地時，我們略過中途的轉機地。例如：以荷蘭航空航班 KL877 為例，由台灣起飛，會中途停在泰國載客，才再轉飛往英國。
- 7、「航空公司」及「航空代理公司」：此處定義「航空公司」為航班所屬之航空公司，負責班機之劃位、登機及行李托運等一貫化作業。「航空代理公司」為某些外國航空公司考量營運成本及策略，並未實際在台設置上述相關作業手續，多為以代理制度委由國內航空公司代為辦理相關業務，僅執行飛行任務。例如：菲律賓航空班機在台業務，係委由華航代理票務、劃位、登機、行李托運等業務，而飛行任務所需之機長、空服人員則由菲律賓航空執行。

另外，還需注意一點，有數個特定的航班號碼屬於同一個航班，是航空公司為合併載運客人之營運設計，此部份必須加以處理，以避免資料重複。例如：長榮航空代理紐西蘭航空及日本日空航空班機，會將某些特定航班乘客，統一合併載運，即使登機證上之出境班機號碼不同，但實際是仍屬於同一班機。此外，「航空公司」亦為我們規劃的維度屬性，其下計有「航廈別」、「國別」、「城市別」三種。

- 8、「抵台班機號碼」及「離台班機號碼」：此處定義「抵台班機號碼」為從他國機場飛至台灣之班機號碼；「離台班機號碼」為從台灣起飛至他國機場之班機號碼。「班機號碼」格式由兩碼航空公司英文代碼加上三至四碼數字組成。多數班機號碼起降時間為固定時段並飛行同一航線，而少部分航班會定期稍作調整異動。故查緝人員多可從偷渡客搭乘之班機號碼，得知偷渡客登機之時段及偷渡目的地。

3.4.3 資料探勘輸入變數型態及採礦模型角色處理

在修正資料庫內之資料以符合分析目的後，還必須將資料定義出適當的資料型態，以避免在分析實作時產生錯誤。以下我們將輸入資料探勘工具的資料稱為「輸入變數」，而我們在資料探勘的分析工具選擇上，是使用 Microsoft SQL Server 2005。Microsoft SQL Server 2005 對於其所附之各項演算法在處理資料探勘輸入變數的資料類型時，僅可接受 Boolean、Date、Double、Long 及 Text 等 5 種資料型別。以下摘要出「輸入變數」適用的資料型態，如表 2。

而 Microsoft SQL Server 2005 各項演算法採礦模型所需之輸入變數角色，共有 KEY(鍵值)、DISCRETE (類別變數)、CONTINUOUS (連續變數)、DISCRETIZED (離散化變數)、KEY SEQUENCE (序列鍵值)、KEY TIME (時間鍵值) 等 6 種。本研究預計採用的部分，說明如下：

表2 「輸入變數」適用之資料型別

資料型別	適用變數
Date	查獲日期、生日等。
Long	年齡、偷渡人數等。
Text	性別、國籍、班機號碼等。

資料來源：本研究整理

KEY (鍵值)：輸入變數的作用是被系統當作是唯一的值，支援 Date、Double、Long 及 Text 四種資料類型。採礦模型均須有此型態變數，始可進行處理。

DISCRETE (類別變數)：輸入變數之間並無大小、順序之關係，各自是獨立的項目。本研究大部分的變數，均屬此型態，例如：國籍別、航空公司名稱等。

CONTINUOUS (連續變數)：輸入變數之間是有連續數值的特性，有大小之分別，可以互相做加減乘除。支援 Date、Double、Long 等三種資料類型之運算。例如：年齡、犯罪停留時間。

KEY TIME (時間鍵值)：輸入變數的作用是被系統作為標示特定序列的鍵值，支援 Date、Double、Long 及 Text 四種資料類型。例如：時間序列分析區隔的時間值，可以是日、月、年等值。

3.5 軟、硬體設備需求

硬體設備：

- 1、個人電腦 (CPU：Dual Core 2.0 GHZ，RAM：3Gb)

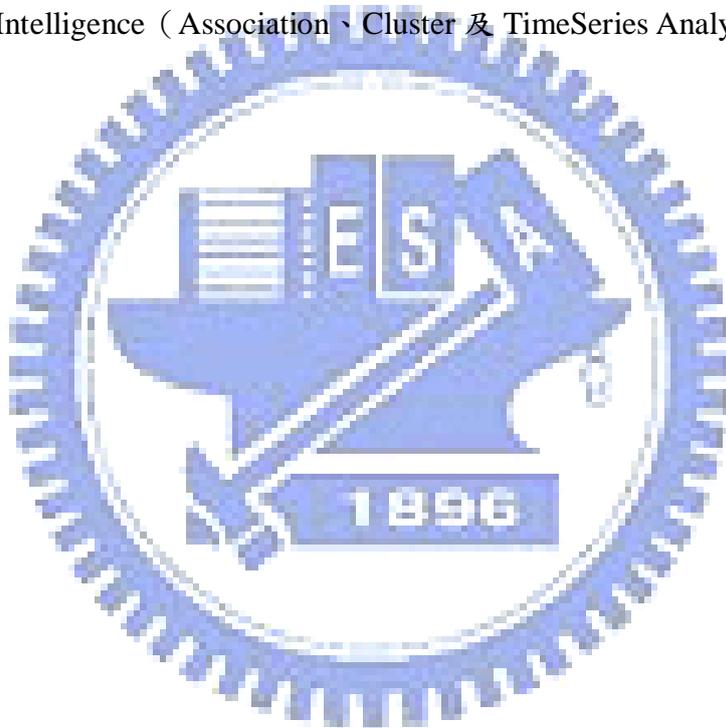
軟體設備：

1、作業系統：Microsoft Windows Server 2003

2、文書處理軟體：Microsoft Office 2003 Professional

3、資料轉換、線上分析工具：Microsoft SQL Server 2005 Management Studio

4、資料探勘及圖表顯示工具工具：Microsoft SQL Server 2005 Analysis Services、
Business Intelligence (Association、Cluster 及 TimeSeries Analysis)



四、系統實作與展示

4.1 建立機場偷渡及班機起降維度資料表

為利因應日後資料擴增之需求、降低系統運算時間，在進行資料探勘的工作前，我們考量先建立資料表的維度分析資料。雖原始資料之資料量並未如一般商業交易資料般繁多，但為考量系統未來之擴展性，建立維度之分析有其必要性存在。預定建立之分析維度如下：偷渡犯罪資料表，以下簡稱為「Snake」資料表。我們預計加入的維度資料表（Dimension）、量值（Measure），如下：

- 1、維度資料表：計有「性別」、「國籍別」、「欲飛往之國別」、「假照國籍別」、「犯罪方式別」、「假照類型別」、「查獲日期」等7個。
- 2、量值：人數（例如：性別、國籍別等）、次數（某年、月、日、查獲次數等）。

民航局桃園機場班機起降資料，以下簡稱為「FlightData」資料表。我們預計加入的維度資料表（Dimension）、量值（Measure），如下：

- 1、維度資料表：計有「航線」、「航空公司」、「載客數」、「目的地機場」等4個。
- 2、量值：架次（例如：班機起降數）、人數（旅客人數）、航線數（航空公司擁有之航線數）。

4.2 Cube 模式分析

建立完上述的維度資料表及量值之後，可從系統中選擇所要察看的 Cube 項目，觀察各項資料的組合、變化情形。此步驟可為經由觀察出圖表的特徵，找出資料探勘的主題，

以提出假設。

4.3 樞紐分析表及圖表建立

在SQL Server 2005 Business Intelligence的「資料來源檢視」功能中，在資料表上選擇「瀏覽資料」即可進入樞紐分析表、圖表功能。SQL Server 2005提供的功能相當強大，只要選擇欲察看的「欄位」，系統即可立即顯示相關的統計圖表，如圖5，圖表之型式可選擇圓餅圖、長條圖、橫條圖等3種。而設計所需要的「樞紐分析表」如圖6，可從右側的「樞紐分析表欄位清單」，以拖曳的方式，將要分析的欄位拖入。使用樞紐分析圖表功能的好處，在於可以在資料探勘作業之前，先行大致判斷出資料的特性及欲分析的方向，不致於在探勘工作進行後，掌握不到探勘的重點。

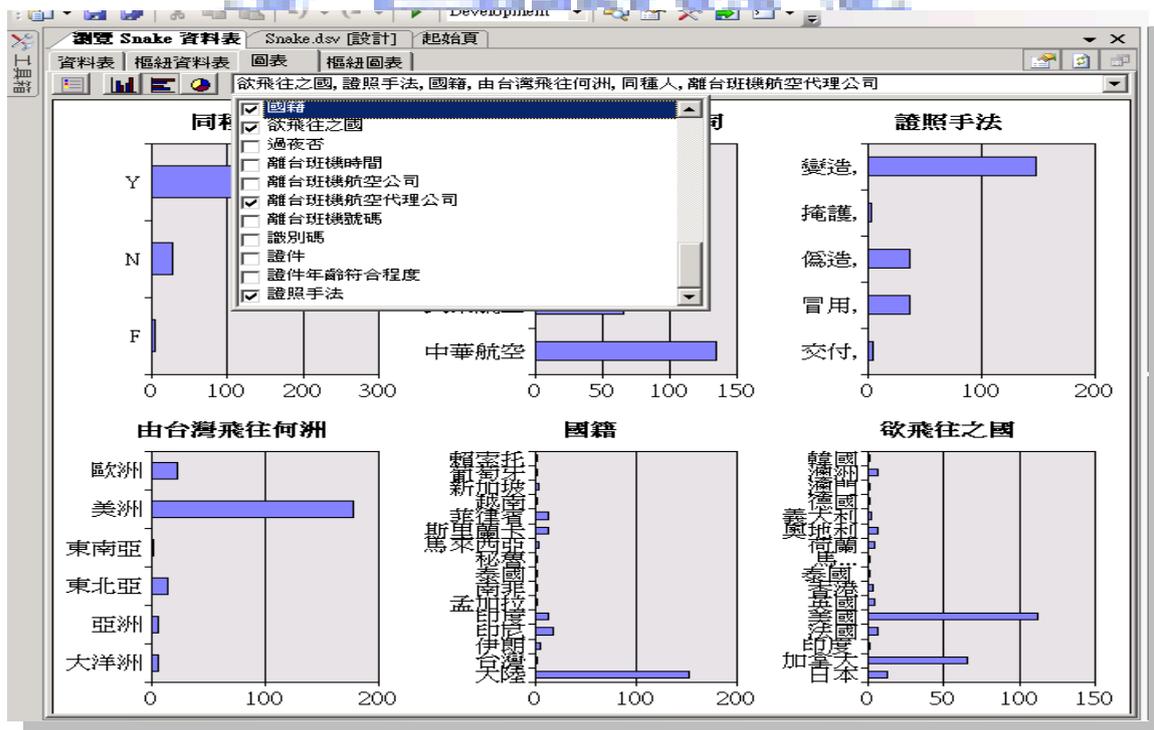


圖5 樞紐分析圖

資料來源：本研究整理

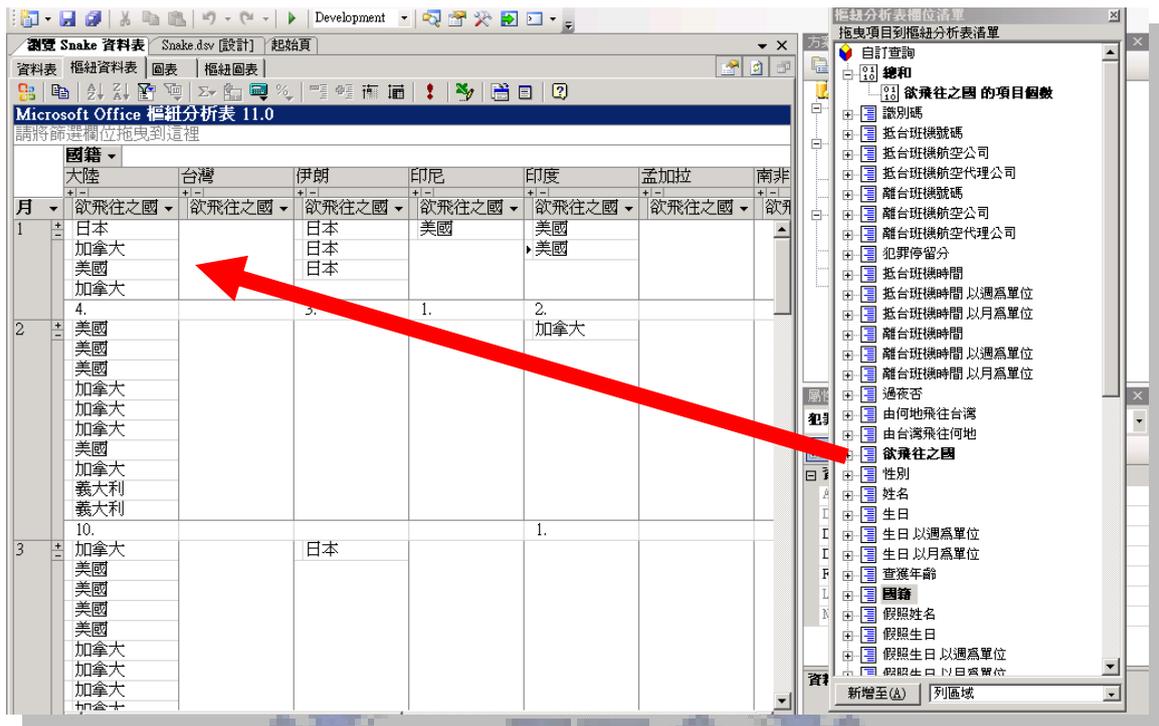


圖6 樞紐分析表

資料來源：本研究整理

4.3.1 國籍與偷渡目的地之樞紐分析

為利資料探勘實施前之準備，我們先針對以下幾個主題來進行樞紐分析：

- 1、性別、國籍、偷渡目的地（Snake 資料表欄位為「由台灣飛往何地」）之關係從圖 7 可以看出，被查獲之偷渡客國籍有 16 個，人數最多前 5 名依序為大陸、印尼、印度，斯里蘭卡、菲律賓。而大陸籍偷渡客前往洛杉磯、及溫哥華之人數最多，佔所有偷渡人數的 24%，高出其他國籍之偷渡客，且其偷渡之目的地很明顯地集中在美、加兩國。而其他國家部份，例如：印尼、斯里蘭卡也有集中於美國之情形存在。在性別的分佈上，皆為男性較多，但大陸及斯里蘭卡籍的男女比例將近相等至各半。

Microsoft Office 樞紐分析表 11.0
請將篩選欄位拖曳到這裡

	國籍												
	大陸	台灣	伊朗	印尼	印度	孟加拉	南非	泰國	秘魯	馬來西亞	斯里蘭卡	菲律賓	越南
由台灣飛往何地	性別	性別	性別	性別	性別	性別	性別	性別	性別	性別	性別	性別	性別
仁川機場/首爾						1.							
巴黎	2.			1.						1.		2.	
布理司班	4.												
札幌	2.												
吉隆坡											1.		
名古屋													
成田機場/東京	5.		4.										
法蘭克福	1.												
阿姆斯特丹												4.	
洛杉磯	55.	1.		6.	3.			1.		2.	11.		
香港	3.												
倫敦	3.											1.	
紐約				3.									
曼谷	1.												
雪梨	2.												
奧克蘭							2.						
溫哥華	52.	1.		3.	9.				1.				
福岡	1.												
維也納												6.	
德里					1.								
澳門													
檀香山	7.												
舊金山	12.			4.							1.		1.
羅馬	2.												

圖7 採國籍—偷渡目的地，採用「性別」偷渡樞紐分析表

資料來源：本研究整理

2、偷渡方式、國籍、偷渡目的地之關係

延續先前的樞紐分析，我們將再探討偷渡之方式（採用從台灣出境或由第三地抵達台灣轉機），對於國籍與偷渡目的地之影響。此處的樞紐分析，考量國籍之偷渡客人數多寡，將範圍限制在最常於台灣進行偷渡的5個國家進行分析。

在機場管制區內進行的偷渡方式有兩種，一種為「入出境」、另一種為「轉機」。「入出境」之旅客都必須逐一經由查驗；「轉機」之旅客則因未入出境，並無需逐一接受身分、證件檢查，而由查緝單位編排人員於旅客登機前隨機執行檢查。故若僅就偷渡之成功率來講，「轉機」的確是可能可以少掉一層被查獲犯罪之機會。而從樞紐分析資料結果如圖8得知，偷渡方式多採「轉機」方式為之，這點相當符合犯罪者之邏輯，且除大陸外，多數國家也相聚台灣較遠，採先行入境台灣再擇日出境的機率較小。

Microsoft Office 樞紐分析表 11.0
請將篩選欄位拖曳到這裡

由台灣飛往何地	國籍				
	大陸	印尼	印度	斯里蘭卡	菲律賓
仁川機場/首爾					
巴黎	2.	1.			2.
布理司班	4.				
札幌	2.				
吉隆坡				1.	
名古屋					
成田機場/東京	5.				
法蘭克福	1.				
阿姆斯特丹					4.
洛杉磯	55.	6.	3.	11.	
香港	3.				
倫敦	3.				1.
紐約		3.			
曼谷	1.				
雪梨	2.				
奧克蘭					
溫哥華	52.	3.	9.		
福岡	1.				
維也納					6.
德里			1.		
澳門					
檀香山	7.				
舊金山	12.	4.		1.	
羅馬	2.				

圖8 採國籍—偷渡目的地，採用「轉機」偷渡樞紐分析表

資料來源：本研究整理

4.4 探勘主題

在進行實際的資料探勘工作之前，我們將查獲到的偷渡客犯罪手法視為人蛇集團的犯罪策略，並據此擬定探勘的主題。這樣的構想主要是針對 2.1.3 節所論及的「空路」偷渡特性來做設計。

本研究將以下列步驟進行探勘。首先，先針對偷渡行為進行「關聯規則」探勘。接著，從前述獲得之資訊，由不同面象透過「分群演算法」分析，取得各群集內所呈現的偷渡活動趨勢。最後，我們將偷渡活動發生的時間週期及加入經濟、失業率指標，以「時間序列演算法」預測偷渡客活動的增減情形，以提供查緝人員於勤務規劃或執行上之輔助參考。

4.4.1 從偷渡方法、目的地，探勘特定國籍偷渡者之喜好度（使用關聯規則）

從前面幾個章節透過樞紐分析圖表可以很快知道數據所呈現的狀態。但透過關聯規則的使用，我們可以瞭解整個偷渡行為彼此之間關聯交錯的情形。在此為求發現何種國籍出發地與目的地之關聯性，我們選擇幾個「輸入變數」做為分析關聯規則的欄位，相關的分析數據設計，如表 3：

表3 關聯規則參數即輸入變數設定表

資料探勘法	關聯規則	
演算方式	Apriori 演算法	
相關參數	最小信心水準>0.8	最小支持度 0.03
變數名稱	變數型態	資料類型
識別碼	Key	Long
性別	Discrete	Text
國籍	Discrete	Text
偷渡方法	Discrete	Text
證件	Discrete	Text
證照手法	Discrete	Text
假照國籍	Discrete	Text
欲飛往之國	Discrete	Text

資料來源：本研究整理

表4 各國籍偷渡客人數分佈表（依照洲別分）

洲名	人數	國籍	洲別	人數	國籍	洲別	人數
亞洲	181	大陸	亞洲	152	印尼	東南亞	17
東南亞	38	印度	亞洲	13	菲律賓	東南亞	13
中東	4	斯里蘭卡	亞洲	13	馬來西亞	東南亞	3
非洲	3	台灣	亞洲	2	新加坡	東南亞	3
美洲	1	孟加拉	亞洲	1	泰國	東南亞	1
歐洲	1		小計	181	越南	東南亞	1
小計	228	伊朗	中東	4		小計	38
		南非	非洲	2	秘魯	美洲	1
		賴索托	非洲	1	葡萄牙	歐洲	1

資料來源：本研究整理

因為資料的母體數並不多，所以在解讀資料時為求符合分析標的，我們會先將整個項目集（Itemset）之規則過濾條件逐步過濾，以求，設定為，「最小支持度」（「支持度」在 SQL Server 2005 中翻譯為「支援」）設定為 24、「項目集大小下限」設定為 3，而從表 4 看來，來自亞洲偷渡國家前 3 名之大陸、印度、斯里蘭卡等國籍之偷渡旅客，佔所有人數 8 成之多。

我們先從規則集內尋找有意義的規則，再對規則所顯示的意義進行闡述，相關結果於圖 9、10 所示。為利將規則集所小，我們先將「信心水準」（「信心水準」在 SQL Server 2005 中翻譯為「機率」）定為 0.8 以上，但若「信心水準」=1.0 且人數少於 5 名的，則不予解釋此條規則，這是因為考量到偷渡案例過少，分析上不具代表性。

支援	大小	項目集
91	3	證照手法 = 變造,, 國籍 = 大陸, 證件 = 護照,
91	3	國籍 = 大陸, 偷渡方法 = 轉機, 證件 = 護照,
71	3	證照手法 = 變造,, 國籍 = 大陸, 偷渡方法 = 轉機
63	3	性別 = 男, 國籍 = 大陸, 證件 = 護照,
59	3	欲飛往之國 = 美國, 國籍 = 大陸, 偷渡方法 = 轉機
58	3	欲飛往之國 = 美國, 國籍 = 大陸, 證件 = 護照,
56	3	性別 = 男, 國籍 = 大陸, 偷渡方法 = 轉機
55	3	性別 = 女, 國籍 = 大陸, 偷渡方法 = 轉機
55	3	性別 = 女, 國籍 = 大陸, 證件 = 護照,
52	3	性別 = 男, 證照手法 = 變造,, 國籍 = 大陸
51	3	性別 = 女, 欲飛往之國 = 美國, 國籍 = 大陸
43	3	欲飛往之國 = 加拿大, 國籍 = 大陸, 證件 = 護照,
41	3	性別 = 女, 證照手法 = 變造,, 國籍 = 大陸
40	3	欲飛往之國 = 美國, 證照手法 = 變造,, 國籍 = 大陸
37	3	欲飛往之國 = 加拿大, 證照手法 = 變造,, 國籍 = 大陸
36	3	欲飛往之國 = 加拿大, 國籍 = 大陸, 偷渡方法 = 轉機
34	3	欲飛往之國 = 加拿大, 性別 = 男, 國籍 = 大陸
34	3	假照國籍 = 新加坡, 國籍 = 大陸, 證件 = 護照,
31	3	假照國籍 = 新加坡, 國籍 = 大陸, 偷渡方法 = 轉機
31	3	假照國籍 = 新加坡, 欲飛往之國 = 美國, 國籍 = 大陸
28	3	假照國籍 = 台灣, 國籍 = 大陸, 證件 = 護照,
27	3	偷渡方法 = 出境, 國籍 = 大陸, 證件 = 護照,
24	3	假照國籍 = 新加坡, 性別 = 女, 國籍 = 大陸

圖9 關聯規則：支持度與項目集表

資料來源：本研究整理

機率	重要性	規則
0.882	2.254	國籍 = 印尼 -> 假照國籍 = 印尼
0.846	2.240	國籍 = 菲律賓 -> 假照國籍 = 菲律賓
1.000	2.233	國籍 = 南非, 證件 = 護照, -> 假照國籍 = 南非
1.000	2.233	國籍 = 南非, 偷渡方法 = 轉機 -> 假照國籍 = 南非
1.000	2.233	國籍 = 南非 -> 假照國籍 = 南非
1.000	2.233	國籍 = 南非, 證照手法 = 變造, -> 假照國籍 = 南非
1.000	2.233	國籍 = 南非, 性別 = 男 -> 假照國籍 = 南非
1.000	2.184	國籍 = 秘魯, 證件 = 護照, -> 假照國籍 = 墨西哥
1.000	2.184	國籍 = 賴索托, 偷渡方法 = 轉機 -> 假照國籍 = 賴索托
1.000	2.184	國籍 = 孟加拉, 證件 = 護照, -> 假照國籍 = 孟加拉
1.000	2.184	國籍 = 孟加拉, 性別 = 男 -> 假照國籍 = 孟加拉
1.000	2.184	國籍 = 秘魯 -> 假照國籍 = 墨西哥
1.000	2.184	國籍 = 孟加拉 -> 假照國籍 = 孟加拉
1.000	2.184	國籍 = 秘魯, 偷渡方法 = 轉機 -> 假照國籍 = 墨西哥
1.000	2.184	國籍 = 賴索托, 證件 = 護照, -> 假照國籍 = 賴索托
1.000	2.184	國籍 = 賴索托 -> 假照國籍 = 賴索托

圖10 關聯規則：「規則表」

資料來源：本研究整理

3 過濾出有意義之關聯規則如下：

我們過濾的規則的優先順序為「支持度」→「信心水準」→「最低重要性」。此乃考量規則內所含的「支持度」過小可能造成「信心水準」較高的假象，而「最低重要性」的值愈高，則代表愈重要。

- (1)、**印尼籍偷渡客偏好使用同自己國籍之護照，八成以上為男性，且偷渡方法採「出境」、「轉機」幾乎各半的機率，偷渡到美國。**

一般偷渡客因出發地距離台灣較為遙遠，除大陸籍偷渡客可就近以「海路」入境台灣外再伺機出境外，多數國家的機場偷渡，應多以「轉機」為主。而印尼籍偷渡客其國家距離台灣有段距離，選擇以「出境」的偷渡方法比率竟達到百分之五十左右，此規則可協助查緝人員注意這個現象，在查緝印尼籍旅客入境時，不可當作一般以勞務為目的之外國人方式查驗，忽略其入境之真正目的。

- (2)、**菲律賓籍偷渡客偏好使用同自己國籍之護照，進行偷渡行為。**

多數的經濟較不佳之國家，都不會願意以自己同國籍的證件來進行偷渡，以避免遭目的地國家在入關查驗時以無正當理由入境或恐逾期停留刁難。而(1)、(2)兩條規則，讓我們發現來自東南亞偷渡國家前2名之印尼、菲律賓國籍偷渡客，同是使用與自己國籍相同之假護照。但偷渡目的地國家卻完全不同，菲律賓國籍是偏歐洲的奧地利（5人）、荷蘭（3人）、法國（2人）、英國（1人），與印尼國籍的美、加顯然不同。

針對上述規則可以協助查緝人員在查驗檢查印、菲籍旅客時，應特別注意核對此類國籍護照之基本資料頁及相關防偽功能，不可在外貌年齡及語言看起來相符，就漏忘檢查細節資料。一般而言，查緝人員僅會先針對護照之真偽、基本資料頁查證。此時應該多加檢查「簽證」、「登機證」是否為飛往特定國家之班機。而印尼籍偷渡客

之目的國以美洲國家為主，菲律賓國籍常偷渡的國家幾乎均屬於歐洲國家，這種情形也讓我們瞭解，不可將偷渡客所選擇的「離台班機抵達地」視為偷渡的最終站「目的地」。因為通常進行「空路」的方式偷渡，都不會是直達偷渡的目的地，後續可能仍會搭配「陸路」或「海路」的方式迂迴抵達真正的目的地。故我們研判菲律賓籍偷渡客的偷渡目的國較為分散，有可能是採用迂迴方式抵達真正的目的地。

4.4.2 以關聯規則分析偷渡喜好（使用關聯規則）

多個「輸入條件」的關聯結果，我們只能得到局部的規則，但還不足以提供大多數資料所可以呈現的關聯規則來理解機場偷渡行為。接下來我們將個別針對偷渡客對「航班號碼」、「航空公司」、「航空代理公司」、「機場」、「國家」、「假照國籍」等幾個主題進行關聯規則的分析，以探討偷渡客之喜好度。

這裡我們將使用關聯規則以常見的購物籃應用分析角度，來切入幾個我們想要分析的主題，而所採取分析的理由，請參考表 5 所示：

表5 以關聯規則分析偷渡喜好之理由

關聯規則主題	意義	分析理由
航班號碼	探討偷渡客對於選擇「抵台航班」、「離台航班」之關聯分析	排除其他因素，僅從中探討偷渡客、人蛇集團對於特定「航班號碼」的組合，選擇之喜好。
航空公司	探討偷渡客對於選擇「抵台航空公司」、「離台航空公司」之關聯分析	除部分國外航空外，航空公司並不會在台灣僅飛單一航線。例如：華航有飛美、歐洲等多條航線。而同一航線及所搭配的機場，亦會有多家航空公司來飛。分析時，我們先排除因航班時刻表之安排差異，僅就「航空公司」這個項目，來探討偷渡客、人蛇集團的喜好。

關聯規則主題	意義	分析理由
航空代理公司	探討偷渡客對於選擇「抵台航空代理公司」、「離台航空代理公司」之關聯	某些外國航空公司為了某些經營上的考量，並不會實際在台灣成立辦理票務、劃位、地勤等部門，而是採用代理之制度委託當地航空公司辦理相關業務，對此我們將探討是否有較多的偷渡行為是集中在某些特定的「航空代理公司」。
機場	探討偷渡客對於選擇「從何機場來台」、「離台至何機場」之關聯分析	領土較為遼闊的國家擁有一個以上之國籍機場，若偷渡客、人蛇集團特別喜好選擇某項機場組合偷渡，從中我們或可取得人蛇集團的運作模式。
國家	探討偷渡客對於選擇「從何國來台」、「離台至何國」之關聯分析	一個國家的經濟好壞，決定偷渡客偷渡地點絕大部分之因素。而特定「國籍」偷渡客對於特定「國家」的喜好，我們或可從中得知人蛇集團的安排此路線之動機為何？
假照國籍	探討偷渡客對於選擇「假照國籍」之關聯分析	偷渡客在從事偷渡行為時所選擇採用之手法並不一致，而我們也好奇，其所持用的護照若係偽、變造，那麼人蛇集團對於假照之國籍身分又如何安排？我們將從此處分析探討。

資料來源：本研究整理

此處採用的關聯規則設計，是將購物籃行為套用於偷渡客的行為上。一般而言，於量販店消費的顧客，其每次消費的商品數量都不會一樣，導致在分析時所得到的頻繁項目集 (Frequent Itemset) 大小也不會固定在某個數量。例如：某顧客 A，第一次消費買了三樣不同的物品，第二次消費買了五樣物品，最後分析出的頻繁項目集或許為三個。而我們這裡所採用的方式因資料特性較為不同，故設計上是將偷渡客在台灣所採取的偷渡行為（例如：搭機來台之行為、搭機離台之行為）視為一筆交易，而每次交易的項目都為 2 個。

除了上述的架構設計外，我們在實作時上述 6 個主題時還必須注意一點，除了抵台與離台的「航班號碼」並不會重複之外，其他分析主題我們都必須將「抵台」與「離台」的值加以處理，讓系統視為不同的項目。以「航空公司」關聯分析為例，搭乘「華航」班機來台，又搭乘「華航」班機離台，我們會在「離台」的部分，後方加註 OUT，以供識別。若未加以處理，系統就會將此筆交易自動辨識為一個物品數量為 2，而從關聯規則中刪除。

為利後續的分析設計，我們以 EXCEL VBA 為工具撰寫程式碼轉換原先的資料庫為合適的架構。然後再依序針對「航班號碼」、「航空公司」、「航空代理公司」、「機場」、「國家」、「假照國籍」這幾個主題進行關聯規則的分析。相關的參數設定，統一設定如表 6。後續的分析結果為節省圖例數量，我們將直接針對執行關聯規則分析後所產生之「相依性網路」圖解釋分析之結果。補充一點，在分析時除「相依性網路」外，「項目表」、「規則」等數據仍應加入輔助判讀，分析結果會更趨於客觀。

表6 以購物籃分析為主題之關聯規則參數設定

資料探勘法	關聯規則	
演算方式	Apriori 演算法	
相關參數	機率>0.4	最小支持度 0.03

資料來源：本研究整理

4.4.2.1 「航班號碼」關聯規則分析

由台灣飛往特定機場的航班，從每日一班甚至數班到每週幾一班到數班不等。固定且每日飛行的航班，相對提供了較多且穩定的機會給人蛇集團從事偷渡行為。此部分的分析結果顯示了偷渡航班間的關聯情形。圖例中的數字代表各節點中心，周遭節點與其關聯的強度。而數字愈小，代表關聯強度愈強。關聯強度之數字僅用來識別節點中心與周遭哪一個節點的關聯規則強度，並非代表符合條件的案例數也成正比。

以圖 11 為例，若未特別標示數字區別關聯強度，代表強度均等。藍色圓形區塊所圍住之 CI032、BR010 為飛往加拿大溫哥華機場之航班。以 CI032 航班為中心來看，由 CI604、CI602 航班連結過去的強度最強，而 CI632、CI650、BR076 等航班連結過去的次之。紅色圓形區塊所圍住之 CI006、CI008、SQ030、MH094、BR016 為飛往美國洛杉磯機場之航班。綠色圓形區塊所圍住之 CI004、BR018 為飛往美國舊金山機場之航班。以 CI004 航班為中心來看，由 CI636、CI614 航班連結過去的強度最強，而 CI674、CI642 等航班連結過去的次之。而我們將較有意義的規則依照「支持度」→「信心水準」→「最低重要性」的順序整理列出如表 7 所示。

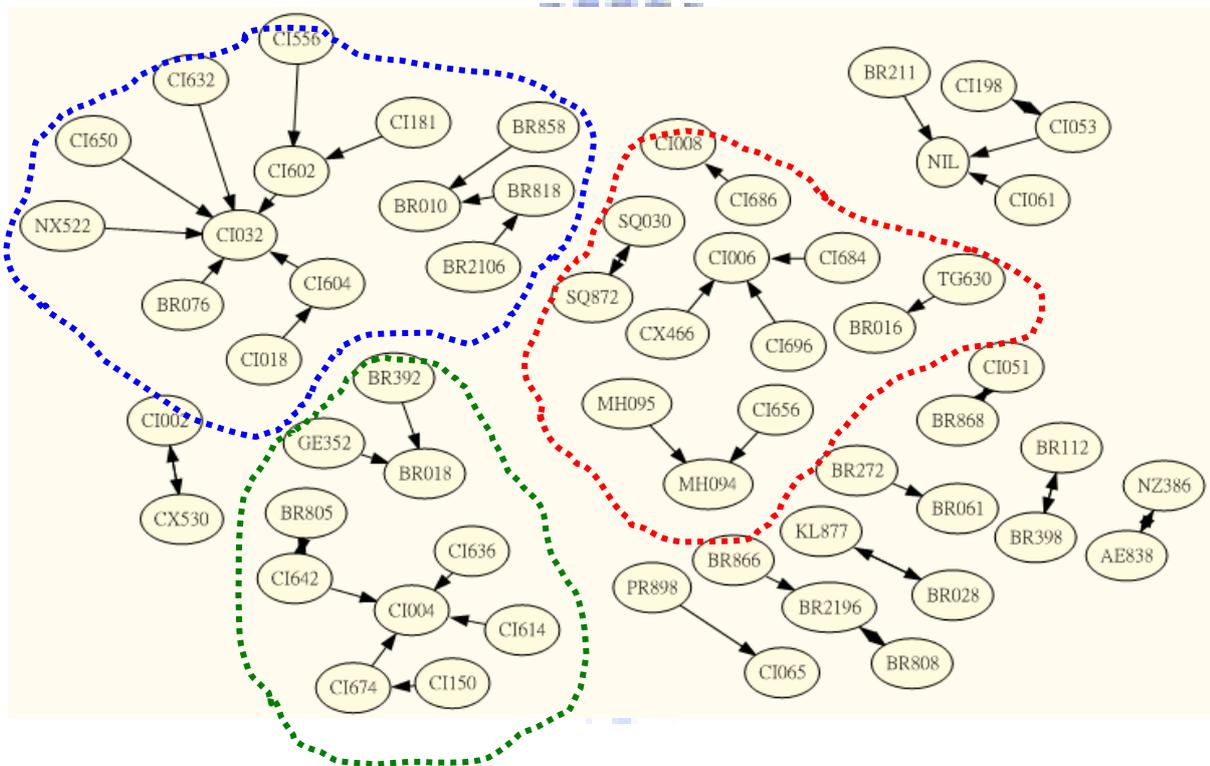


圖 11 「航班號碼」關聯規則分析結果「相依性網路」圖

資料來源：本研究整理

一個特定航班，擁有愈多的連結，除了代表多個抵台的航班會往此航班移動，進而讓我們去瞭解航班與航班間之喜好度。其作用可以提供查緝人員擬定查緝策略於特定航班組合的到、離台登機門沿線，尋找於機場管制區遊移的偷渡客。

另一方面，從航班所屬之航空公司來看，我們發現大部分的抵台、離台航班組合所屬的航空公司，均屬搭配同一公司為多，故從分析結果可以得知抵台及離台航班屬於同一個公司或代理公司會增加人蛇集團喜好度，且若抵台及離台航班又屬於同一航廈登機，也相同會增加人蛇集團的喜好度。

表7 「航班號碼」關聯規則整理表

信心水準	重要性	「航班號碼」規則	支持度
0.75	0.910331	CI662 -> CI006 (飛往洛杉磯)	12
1.00	1.156872	SQ872 -> SQ030 (飛往洛杉磯)	9
1.00	0.756527	TG634 -> CI032 (飛往溫哥華)	8
0.73	1.004652	BR228 -> BR010 (飛往溫哥華)	8
0.78	0.825878	CI606 -> CI006 (飛往洛杉磯)	7
0.55	1.469539	CI678 -> CI008 (飛往洛杉磯)	6
0.50	0.806978	BR818 -> BR010 (飛往溫哥華)	6
1.00	1.285236	CX402 -> CI004 (飛往舊金山)	5
0.71	0.59243	CI602 -> CI032 (飛往溫哥華)	5

資料來源：本研究整理

4.4.2.2 「航空代理公司」關聯規則分析

於前面第三章我們提到，某些外國航空公司考量營運成本及策略，並未實際在台設置上述相關作業手續，多為以代理制度委由國內航空公司代為辦理相關業務，僅執行飛行任務。在國內，中華航空於國內代理外國航空公司之數量居國內航空之冠。旗下代理的航空公司國際線部分，計有（大馬、荷蘭、華信、菲律賓、越南、澳門航空等），而長榮、復興航空代理之數量，則次之。

由圖 12 看，中華航空與其他航空代理公司的連結線較多且複雜，此種結果與「航班號碼」關聯規則分析圖形差異不大，不過也更容易呈現偷渡客於各主要航空代理公司偷渡的網絡情形。因台灣偷渡行為以飛往美、加、日三國的情形較多，而圖形也呈現集中在主要負責這些航線的三個公司中。而我們將較有意義的規則依序整理列出如表 8 所示。

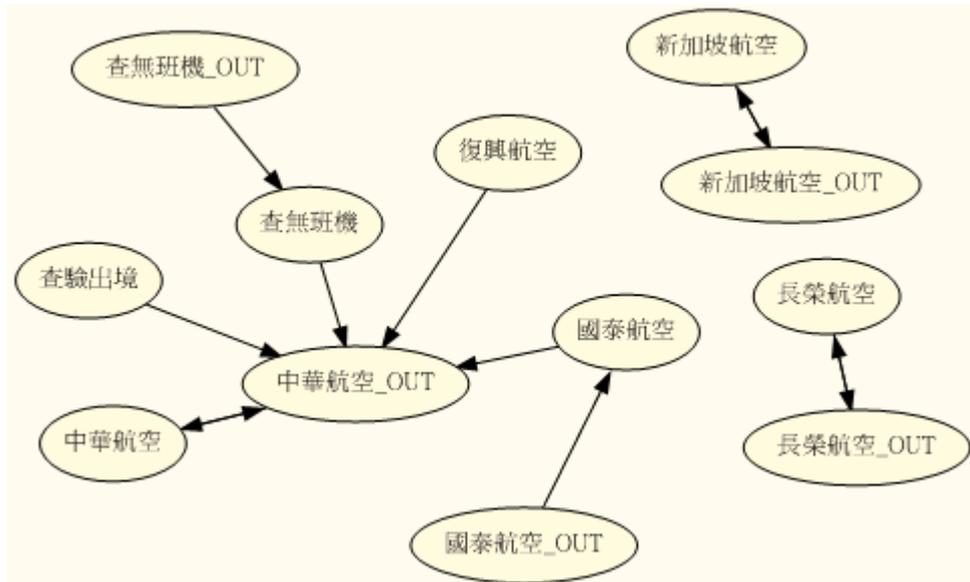


圖12 「航空代理公司」關聯規則分析結果「相依性網路」圖

資料來源：本研究整理

表8 「航空代理公司」關聯規則整理表

信心水準	重要性	「航空代理公司」規則	支持度
0.841	0.358307492	中華航空 -> 中華航空_OUT	90
0.889	0.776493963	長榮航空 -> 長榮航空_OUT	40
1.00	1.403019534	新加坡航空 -> 新加坡航空_OUT	15
0.8	0.126027779	國泰航空 -> 中華航空_OUT	12

資料來源：本研究整理

對於這樣的結果，也讓我們對於人蛇集團是否已買通該特定航空代理公司地勤人員之情形持保留態度[11]。因為多數航空代理公司在為所代理之公司辦理劃位或登機等手續時，考量成本不會額外增設一組地勤人員，而這樣的情形就會導致地勤人員可以在橫跨所代理的多個公司航班作業，故人蛇集團可以藉由買通航空代理公司（例如：中華航空）的地勤人員，即可透過這樣裏應外合的手法，協助偷渡客進行偷渡，此種合作情形尤其在最後地勤人員於登機門前收票作業上更顯得關鍵。

4.4.2.3 「機場」關聯規則分析

一個國家對外的入出境機場，視其國家腹地大小及政治、經濟發達程度，可能有一到數個不等。機場的設施、周邊交通、營運量、合作夥伴等因素也對人蛇集團選擇偷渡行為橫跨之機場影響甚多。由於在機場被查獲之偷渡犯罪者類型中以偷渡客居多，而規劃執行偷渡之人蛇集團成員甚少被查獲，故我們假設人蛇集團的分工仔細而縝密，人蛇集團在整個偷渡流程中，在機場的選擇上也有其專業性。以台灣桃園機場設施為例，監視器遍佈，集團成員如何有經驗、熟練地、技巧性地避開查緝人員，並引領偷渡客前往登機室或交付重要證件，任務失敗之逃亡路線，均需事前經過縝密計畫。人蛇集團對於偷渡行為所牽涉之機場之選擇，固然與台灣和各國所簽訂之航線相關，但我們考量查緝人力資源之有限性，仍有必要進行分析瞭解，以掌握集團的運作手法，進而擬定策略防堵。

以下我們將只針對常被偷渡之目的國機場：美國、加拿大、日本等國，做為說明的標的。從圖 13 中，「機場名」所代表之意義，為「由何機場抵台」。「機場名」後加上「OUT」，在此代表「離台至何機場」。而連結線的箭頭指向位置，反應與形成的案例數量有關，一般多為指向「離台至何機場」，例如：洛杉磯。若反指向「由何機場抵台」，例如：香港，則代表該機場往各地機場的案例多，故其反而成為主角。若某機場有兩個對外指向，則箭頭會指向案例較多的那一個機場，例如：檳城與舊金山、名古屋等機場；若對外的指向案例數相同，則皆朝向「離台至何機場」的機場，例如：德里機場。

而從結果得知美國舊金山機場偷渡客多由檳城機場而來、洛杉磯的偷渡客（依序為由新加坡、雅加達、吉隆坡、胡志明、河內等機場過來）、檀香山多由香港而來。加拿大溫哥華機場偷渡客，依序為由曼谷、胡志明機場過來。日本名古屋機場多由檳城機場而來、成田機場依序為澳門、香港而來。福岡多由澳門而來、札幌多由台北而來。最後，我們將較有意義的規則依序整理列出如表 9 所示。由上述結果，我們不難推測連結線多的機場，可能已有多組國際性的人蛇集團在當地營運。

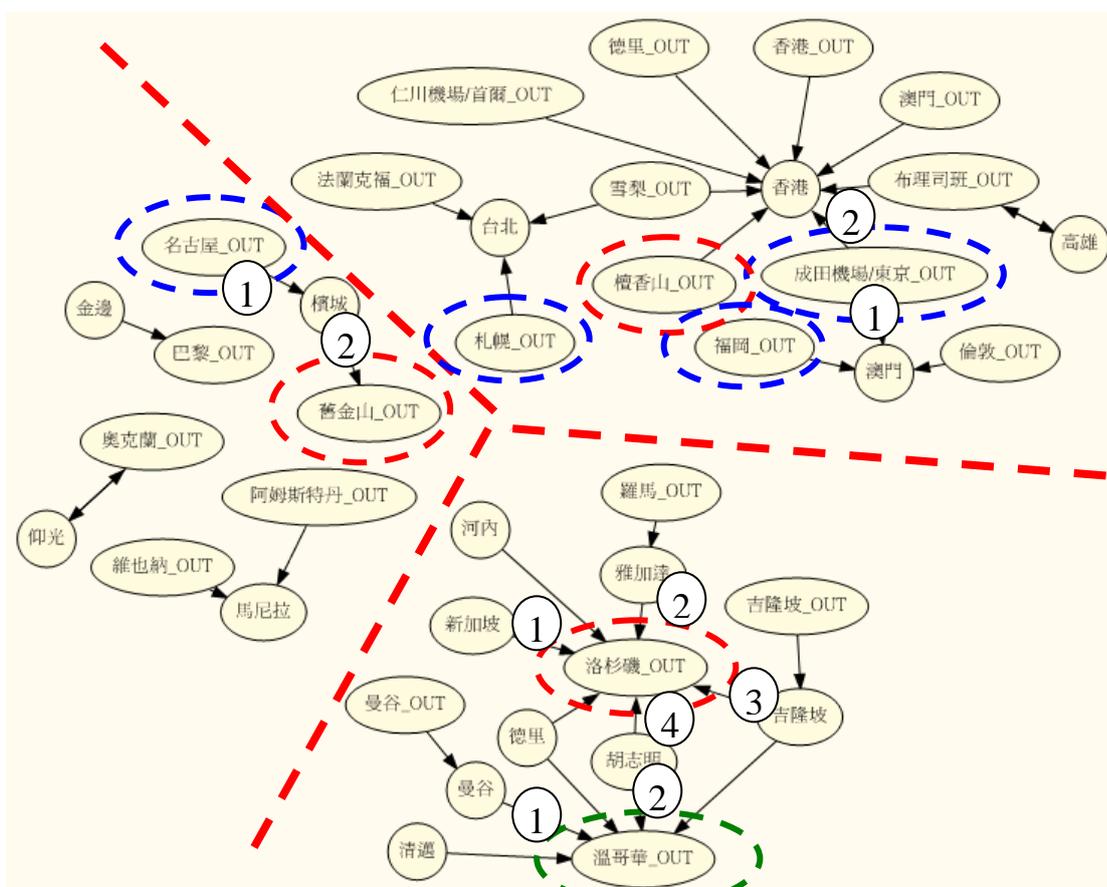


圖13 「機場」關聯規則分析結果「相依性網路」圖

資料來源：本研究整理

表9 「機場」關聯規則整理表

信心水準	重要性	「機場」規則	支持度
0.871	0.479309141	新加坡 -> 洛杉磯_OUT	27
0.565	0.331764613	曼谷 -> 溫哥華_OUT	13
0.526	0.180132986	吉隆坡 -> 洛杉磯_OUT	10
0.421	0.185453658	吉隆坡 -> 溫哥華_OUT	8
1.00	0.526300329	檀香山_OUT -> 香港	7
1.00	1.146128036	維也納_OUT -> 馬尼拉	6
0.545	0.185108077	雅加達 -> 洛杉磯_OUT	6

資料來源：本研究整理

4.4.2.4 「國家」關聯規則分析

圖 14 中，「國名」所代表之意義，為從何國機場抵台。「國名」後加上「OUT」，在此代表「離台至何國機場」。此外，若圖例上有相同國籍互相連結，例如：「香港」→「香港 OUT」。代表從香港，尚未辦理轉機手續，但於掩護或協助他人冒用證件或登機證時被查獲，故我們不將對方當為一般之偷渡客，而當成是協助偷渡之人蛇集團成員。

我們將圖 14 可以分成兩組行為來看偷渡客經由台灣到各國的遷移情形。從圖 14 左側這一組看來。我們發現菲律賓偷渡客多往歐洲國家（荷蘭、奧地利）偷渡，而其官方語言雖為菲律賓語、英語，近半世紀也曾短暫為美國所殖民。但該國早於 1565 年起成為西班牙所殖民地將近 300 多年，所形成之歷史背景、文化，與圖所顯示的遷移歐洲國家現象，乃屬合理之現象。而澳門、柬埔寨在過去曾經分別為英國、法國之殖民地，偷渡英、法國之行為也近乎合理。故我們發現偷渡客對於偷渡國家之選擇上，除了經濟因素外，國家曾被殖民之歷史背景亦為考量之重點。

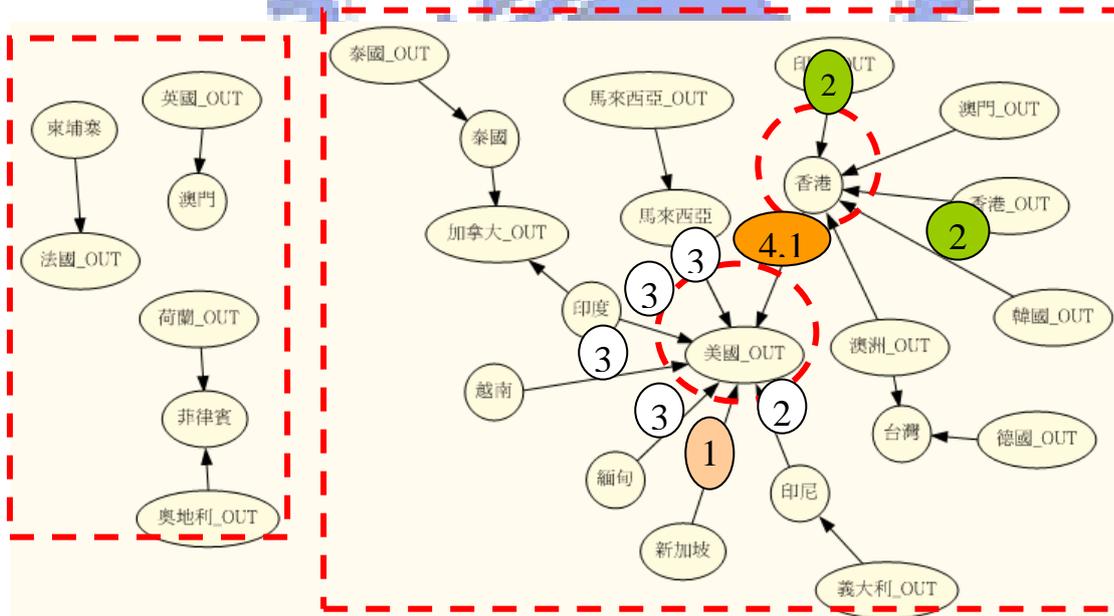


圖 14 「國家」關聯規則分析結果「相依性網路」圖

資料來源：本研究整理

圖 14 右側另一組之行為，則比較複雜多元。我們將「美國」、「香港」各自形成一個群之中心。以「香港」為中心而言，由香港抵台，離台至美、韓、印等國（英文字母愈前面，代表關聯性愈強）比例較高。由此群可知，從香港來台的偷渡客，較有橫跨其他美、亞洲之行為，相較於其他國家遷移情形多僅單一洲之情形有些不同。若以「美國」為中心而言，從新加坡、印尼（以下簡稱「新、印」）抵台，離台至美國之比例相當高。而我們進一步查驗其所持用的假證件資料，發現竟大多也是同出發地之「新、印」護照居多數，我們對於此種現象感到不合理。一般而言，查驗人員應該會對其所屬國家之證件特徵、語言熟練，並能輕易和旅客交談後，瞭解待通關查驗之旅客為國人或外國人。故「新、印」兩國查驗其國人所持假護照之查獲率，應高於其他外國籍證件。

我們排除被查獲偷渡客於出境、轉機偷渡時是使用兩本護照偷渡的機率下，推測這種情形有以下幾種可能：一為由該「新、印」抵台之偷渡客，可能也是屬於不用查驗證件之轉機旅客，故能成功由該國離境。另一種可能為查驗人員之專業能力有待加強或查驗時較為輕忽。但若該國家屬於民主程度落後或經濟發展較弱之國家，我們不排除也有可能是有不肖政府官員與人蛇集團勾結放水。實務上，依筆者在檢查部分印尼護照之經驗，也有此項疑問。

另外，由泰國抵台且離台至加拿大之偷渡行為也佔多數，但其所持用之護照較無一致，無與上述相同之疑慮。雖然在台灣查獲偷渡日本之現象也多，但圖例中卻沒有出現日本的，此乃因為偷渡客遷移之情形較為分散，故不被系統接受形成關聯規則。從上述分析中，我們可以得知偷渡行為的抵台、離台地點是有規則性的，而我們傾向此與人蛇集團所長期營運之偷渡路線有關聯，且單一地區就可能有多數條偷渡路線存在，並非偷渡客可單獨客製化選擇決定偷渡之目的國家。最後，我們將較有意義的規則依序整理列出如表 10 所示。

表10 「國家」關聯規則整理表

信心水準	重要性	「國家」規則	支持度
0.492	0.001438063	香港 -> 美國_OUT	32
0.903	0.313318209	新加坡 -> 美國_OUT	28
0.583	0.350709262	泰國 -> 加拿大_OUT	14
0.583	0.079349937	馬來西亞 -> 美國_OUT	14
0.818	0.209467423	印尼 -> 美國_OUT	9
0.636	0.104284884	越南 -> 美國_OUT	7
1.00	1.146128036	奧地利_OUT -> 菲律賓	6
1.00	1.07080721	荷蘭_OUT -> 菲律賓	4

資料來源：本研究整理

4.4.2.5 「假照國籍」關聯規則分析

圖 15 中，「國名」所代表之意義，為「何國籍之偷渡客」。「國名」後加上「OUT」，在此代表「所持用之假證件國籍」。整張圖我們將「關聯性連結」強度設定到最弱，很容易可以拆成兩組區塊解釋。一組是以「大陸國籍」偷渡客為主，另一組則是東南亞、中亞、非洲國籍偷渡客為主。從圖 16，我們可以發現東南亞（馬來西亞、菲律賓、印尼）、中亞（孟加拉、印度）、非洲（南非、賴所托）等幾個國家都傾向使用偽造或變造與自己國籍相同的證件，來進行偷渡活動。因符合特定規則之比率相當高，故相依性網路工具的圖形也顯示出其為「相互預測」的節點。由於上述所列各國其持用證件的行為都比較一致並無交叉其他國家證件之行為，故不依照「關聯性連結」強度標示號碼以顯示哪一個特定國家其與假照之關聯性特別強。

另外一組以「大陸國籍」偷渡客為主的圖形，其使用的偽變造其他國籍證件的種類就比較具多元性。我們在圖 17 上以號碼標示強度說明，標示的號碼愈小，表示強度愈強。為了簡化說明重點，我們忽略身處周圍未直接與「大陸」國籍連結之國籍，只針對與「大陸國籍」偷渡客所直接連結的國家進行標示。圖 17 所標示的 1-2 號碼看來，「大陸國籍」偷渡客仍以同文同種之國籍證件為最優先考量，其次才是選擇亞洲面孔之國籍且，例如：

至於會持美國與英國海外護照之情形。英國海外護照部分：由於香港在回歸中國大陸之前，其居民持用之護照為「英國海外護照」，而香港鄰接中國且人種相同，故此種狀況我們並不意外。另美國護照部分：我們推測因在美之華人移民已到達某個程度到處可見，尤以有唐人街之城市為盛，若稍懂美國語言、文化或可騙過入境之移民官員。最後，我們將較有意義的規則依序整理列出如表 11 所示。

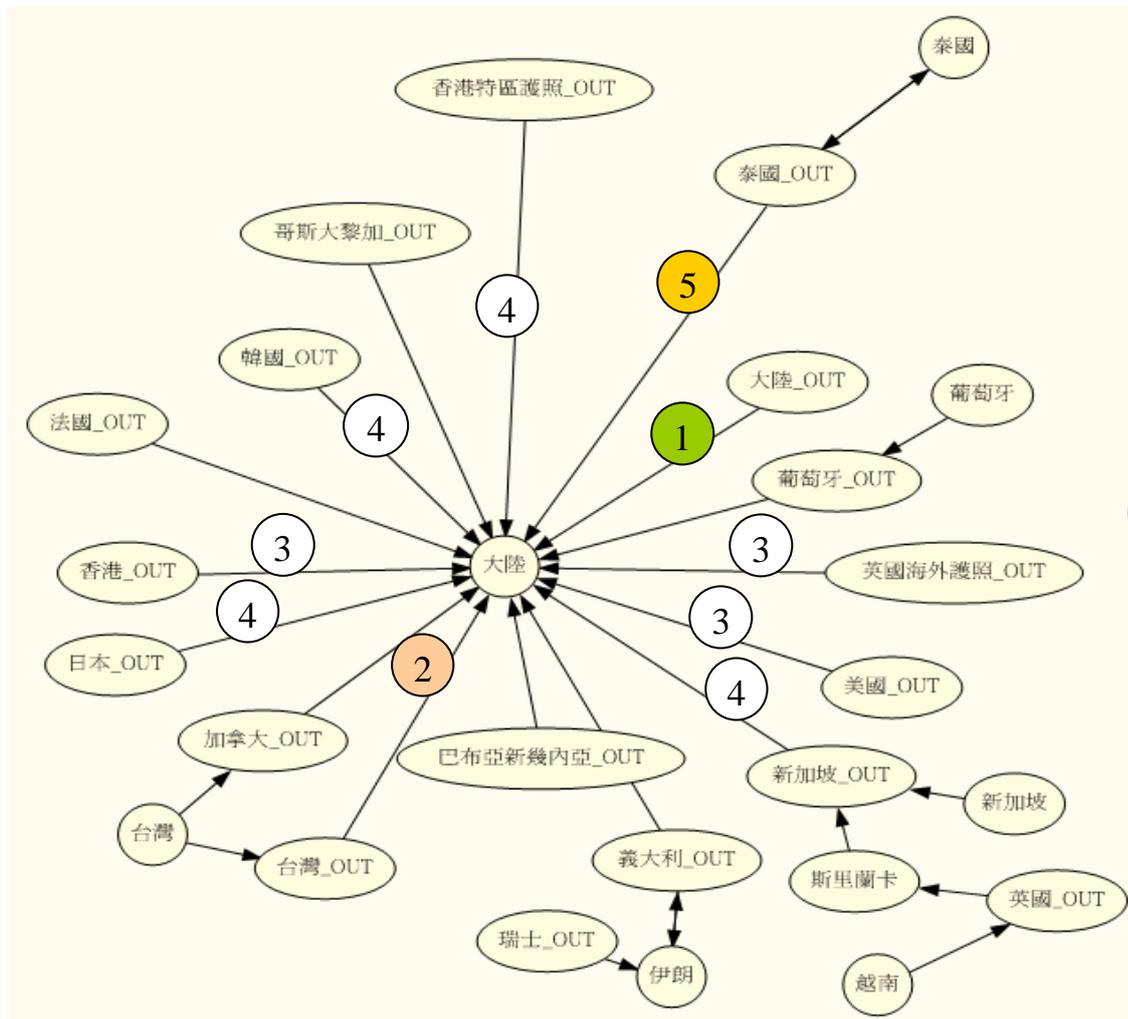


圖 17 「假照國籍」關聯規則分析結果「相依性網路」左側放大圖

資料來源：本研究整理

表11 「假照國籍」關聯規則整理表

信心水準	重要性	「假照國籍」規則	支持度
0.766	0.072270872	新加坡_OUT -> 大陸	36
0.969	0.183976421	台灣_OUT -> 大陸	31
1.00	0.1937181	大陸_OUT -> 大陸	28
0.882	2.253745985	印尼 -> 印尼_OUT	15
0.846	2.239549721	菲律賓 -> 菲律賓_OUT	11
0.769	2.20176116	印度 -> 印度_OUT	10
1.00	0.139227482	美國_OUT -> 大陸	8
1.00	0.139227482	英國海外護照_OUT -> 大陸	8
1.00	0.139227482	香港_OUT -> 大陸	8
0.875	0.085090106	葡萄牙_OUT -> 大陸	7
0.5	-0.1309291	加拿大_OUT -> 大陸	7

資料來源：本研究整理

4.4.3 偷渡客犯罪手法之趨勢（使用群集）

此部分我們將以分群演算法來探討偷渡客進行犯罪行為所採用的手法。為探求出偷渡客的犯罪手法趨勢，我們將現有資料表內相關於犯罪手法的「輸入變數」以分群方法進行分析，藉以瞭解各群之間之相似性如何？以及各群內成員行為所反應出之偷渡趨勢。透過分群演算法，不僅可以找出偷渡客行為間自然的相似結構關係，圖形化介面更可以呈現偷渡客的行為群集特色，這對於我們在判斷人蛇集團的偷渡行為群集有很大的幫助。以下我們將分成兩個部分以分群演算法探討偷渡客之犯罪手法。分析標的其一為偷渡客所持有的證件趨勢；另一為偷渡客選擇目的地的趨勢。

以下我們將針對研究設計部分，歸納為以下幾點詳加說明：

1. 在分群演算法的選擇上，我們採用 E-M 與一般常被使用之 K-means 方法不同。在實務上 K-means 多用於「連續變數」，如果非放「類別變數」不可，也儘可能採用連續性指派(Arbitrarily assignment)的技巧來處理類別變數。本研究資料型態特性因以「類別變數」居多，故不採用一般常見以距離函數計算相似性的 K-means 分群

法，而採以資料點密度作為計算相似性的 E-M 方法；且由於資料母體不大，故 EM 以機率分配的方式計算相似度，可將群集特性的不同以柔性群集的形式細微區分出來。而群數的設定，我們採用系統自動偵測最佳群集，系統所偵測出結果為 4 群。

2. 在「輸入變數」的選擇上，我們在研究設計中不將偷渡者之「年齡」列入，此乃考量會偷渡之勞動人口年齡（本資料人數年齡分佈以 15-45 歲居多），對於我們在判斷犯罪手法上並無太大助益，且以現今時代年齡與外表之差距辨識上已不明顯，故捨棄不用。
3. 另本研究「輸入變數」的設計與前述之「關聯規則」多有相似，但我們額外加入了「犯罪停留時間」，因關聯規則並無法將「連續變數」的部分處理地很好，故我們透過分群的方式來處理。另考量採用分群演算法時，「輸入變數」儘量避免意義重複之變數。而一般之看法會認為「犯罪停留之時間」與「在機場過夜否」之相關性成正比，但實際上仍有一段差距，並非全然成正比。「犯罪停留之時間」長短，牽涉到下一段行程班機的時刻安排，以台灣桃園機場預定之客機班機時刻表（不含貨機在內）為例，每日入出境航班之起迄時段約為清晨五點多至凌晨十二點多左右。機場客機停止起降之空檔時間約為五個小時左右內。分析「在機場過夜否」，可提供查緝人員瞭解對於無班機時段的機場管制區清場檢查或過境旅館執行旅客之身分清查，執行的效益有多少？另於從前述的航空公司關聯分析中我們得知人蛇集團在安排偷渡，多數會將抵台、離台班機之安排，傾向選擇同一公司，故我們在變數設計時，傾向不將「犯罪停留之時間」與「在機場過夜否」所代表之意義，視為重複。

1、偷渡客犯罪手法之趨勢（分群分析）分析設定數據，如表 12：

表12 偷渡客犯罪手法之趨勢（分群參數設定）

資料探勘法	分群演算法	
演算方式	E-M 演算法（以資料點密度為分群方法）	
相關參數	群數=4（自動偵測之結果）	機率聚合門檻:0.1
變數名稱	變數型態	資料類型
識別碼	Key	Long
國籍	Discrete	Text
欲飛往之國	Discrete	Text
假照國籍	Discrete	Text
犯罪停留分	Continuous	Double
在機場過夜否	Discrete	Text
偷渡方法	Discrete	Text
證件	Discrete	Text
證照手法	Discrete	Text

資料來源：本研究整理

2、「解讀分析與建議」：

從分析結果的整體來看，「群1」有104例、「群2」有62例、「群3」有38例、「群4」有24例。各群的相似性如圖18所示，最高者是「群3」與「群4」，而「群3」與「群1」次之。若以群集的特性來看，如圖19所示「群3」與「群4」兩群之間內的國籍組成份子較為複雜，與「群1」、「群2」主要由大陸國籍組成的不同。

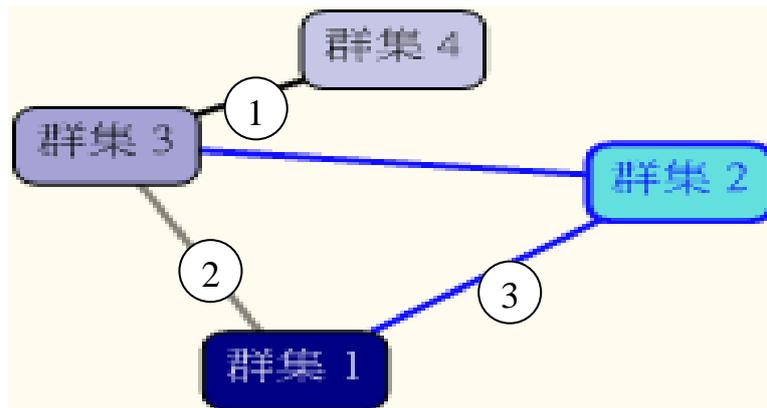


圖18 偷渡犯罪手法群集表，各群間連結強度（強度愈強，數字愈小）

資料來源：本研究整理

若個別就「犯罪停留之時間」與「在機場過夜否」配合「國籍」來看，群1、群2的偷渡效率是比較高的，幾乎不在機場過夜。而「國籍」內的成員配合機場班機時刻表來看，我們可以理解因為香港來台之航班數量較多且多數大陸偷渡客均由香港來台轉機，故航班之接駁配合上可以較為緊湊。而其他群內外國籍的偷渡客因抵台航班接駁班次之故，普遍需要較長時間之等待。

從「國籍」來看，群1與群2雖然近似，但其偷渡持用的證件種類與證件偽變造手法，卻大大不同，據此，我們懷疑此兩群有可能屬於不同的人蛇集團所安排。

「群集4」與「群1-3」的偷渡手法較為不同，以轉機為主。觀察各群之內「國籍」屬性的組成結構如圖20所示，以菲律賓籍居多、印度等次之。我們可以得知以多數大陸籍為主之偷渡客，選擇偷渡之方法較為多元，因從大陸至台灣之距離較近，且台灣四面環海，海路偷渡發達，大陸籍偷渡客可分段先以海路入境台灣打工或停留，再擇日出境。而其他國籍偷渡客相距台灣較遠，偷渡難度成本較高，也無必要花費太多成本採行先入境台灣再由台灣出境之方式遂行偷渡。另欲偷渡前往之目的地國家，也非如其他群集以美、加等國為主。



圖19 偷渡犯罪手法趨勢「群集設定檔」

資料來源：本研究整理



圖20 偷渡犯罪手法「群集辨識」

資料來源：本研究整理

4.4.4 偷渡客選擇目的地機場的趨勢（使用群集）

在此部分研究設計中，我們欲探討的重點在於偷渡客選擇目的地的行為。而為與前述關聯規則分析有所區別不重複，我們選擇偷渡行為涉及之「機場」，而非「國家」。我們所考量的是一個國家依其地域大小分設不只一個以上之機場，而機場工作人員之調動並不頻繁，若人蛇集團要買通相通移民官員或航空公司人員等建立關係，必定會採取長久之合作模式，而其為獲利、提高犯罪成功率，也會儘量縮短犯罪停留之時間。

1、偷渡客選擇目的地的趨勢（分群分析）分析設定數據，如表 13：

表13 偷渡客選擇目的地的趨勢（分群參數設定）

資料探勘法	分群	
演算方式	E-M 演算法（以機率函數為分群）	
相關參數	群數=3（自動偵測之結果）	機率聚合門檻:0.1
變數名稱	變數型態	資料類型
識別碼	Key	Long
國籍	Discrete	Text
犯罪停留分	Continuous	Double
由台灣飛往何地	Discrete	Text
由何地飛往台灣	Discrete	Text

資料來源：本研究整理

從分析結果的整體來看，各群的相似性最高者是「群1」與「群2」，而「群3」獨立一群，如圖 21 所示。而「群1」有 104 例、「群2」有 72 例、「群3」有 52 例。如圖 22 所示。

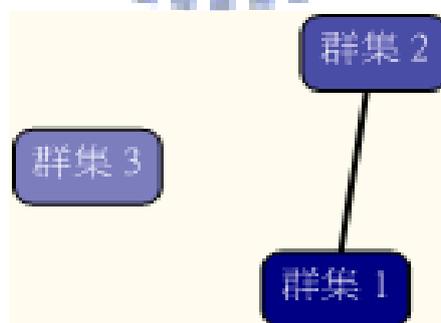


圖21 偷渡客選擇目的地的趨勢「群集辨識」

資料來源：本研究整理

2、「解讀分析與建議」:

以群集的特性來看，「群 1」，國籍以大陸偷渡客為主，選擇多由鄰近大陸的香港機場抵達台灣進行偷渡，偷渡之目的地點以加拿大的溫哥華機場為主，且在整個過程中平均耗時，換算單位（127.68～ 225.69 分鐘）為 2-6 個小時多。

「群 2」，國籍以大陸、印尼偷渡客為主，選擇不特定之機場抵達台灣進行偷渡，偷渡之目的地點以美國的洛杉磯機場為主，且在整個過程中平均耗時相較「群 1」為久，換算單位（156.59～ 241.69 分鐘）為 3-6 個小時多。

「群 3」，獨立自成一群，群內國籍不特定，選擇不特定之機場抵達台灣進行偷渡，偷渡之目的地點雖亦有美、加等國機場，但其他國家之機場亦不在少數。而且在整個過程中平均耗時為三個群集中最久，換算單位（666.52～ 1087.73 分鐘）為 11-18 個小時多。

從上述兩個分群的結果，我們可以得知人蛇集團安排大陸籍偷渡客偷渡時，因佔地理位置、航班時段之優勢，故進行偷渡較有效率，選擇「海路」、「空路」之搭配也較有彈性。而其他國籍利用台灣做為偷渡之跳板，除可能一開始是因為與台灣開放政策引進外勞，例如：印尼、泰國、菲律賓等國籍勞工，或與外籍人士大量通婚，而將此路線打開，另一方面也由於台灣於國際上之形象良好，故使得台灣亦成為上述國家轉機偷渡喜好之點。而此類國家偷渡客選擇偷渡目的地，因有其語言、文化之差異，亦多與大陸籍旅客南轅北轍。不過從整體來看，各國偷渡客採行偷渡之手法、路線不同，也難令人察覺出各國人蛇集團在台有跨國性之合作偷渡情形。



圖22 偷渡客選擇目的地的趨勢「群集設定檔」

資料來源：本研究整理

4.4.5 偷渡客人數、時段之影響預測（使用時間序列分析）

4.4.5.1 偷渡客人數、時段之趨勢預測

在得知偷渡客的行為之後，我們從資料中可以得知偷渡的行為的發生，有所謂的熱門週期，以機場客機時刻預定表來看，每日固定對飛的班機，為何偷渡行為總是集中在某特定時段？星期幾？或某個月份？這個部分則有待進行實驗以證明我們的假設。在此部分研究中我們將僅針對偷渡客最喜好的3個國家（美國、加拿大、日本）進行處理，藉以預測出未來的偷渡人數增減趨勢。

接著，我們為了能夠事前瞭解偷渡的趨勢，我們蒐集了特定國家偷渡人數與該國的經濟數據，研究其中是否有相對應的關係，藉以輔助判斷偷渡趨勢提出警示。這個部分的研究

究由於蒐集到的資料量週期不大，所預測出之趨勢結果有待評估，故我們將以分析結果所呈現之圖形做為主要分析的判斷，而不以時間序列法得到的迴歸係數結果做為分析標的。相關的分析數據設計如表 14：

表14 偷渡客人數、時段之時間序列分析參數設定

資料探勘法	時間序列	
演算方式	自我迴歸樹演算法 (ART)	
相關參數	機率>0.8	最小支持度 3
變數名稱	變數型態	資料類型
時間區塊	Key Time	Date
偷渡目的國	Key	Text
偷渡人數	Predict	Long

資料來源：本研究整理

此處為了能夠顯示出資料預測的「季節性」部分，我們額外參數「週期提示」Periodicity_Hint 輸入{12}。分析的結果如圖 23、24，由圖我們可以得知由台灣轉偷渡美、加的人數，高峰期多在台灣適逢暑假的這幾個月，而其他國家偷渡客也有此類似情形，據此可推測人蛇集團於此期間台灣機場旅客運量高峰時採取偷渡行為，乃為藉機蒙混過關。而有趣的是偷渡美、加的人數高峰趨勢情形大多一致，如圖 24 所示。而圖 23、24 虛線部分所呈現的是未來 5 期（此處單位為月份）的預測值。以美國偷渡人數的當期預測為例， $= 1.715 + 6.548 * \text{數量}(\text{荷蘭}, -1) + 2.117 * \text{數量}(\text{日本}, -1) + 3.159 * \text{數量}(\text{法國}, -1)$ 。而我們對於每週的偷渡人數分佈情形也會有興趣，此處將參數 Periodicity_Hint 設定為{7}。從圖 25 可以看出偷渡人數不分偷渡目的地，偷渡尖峰集中週四至週日這段期間，而隨著回到星期一時，人數逐漸下降。

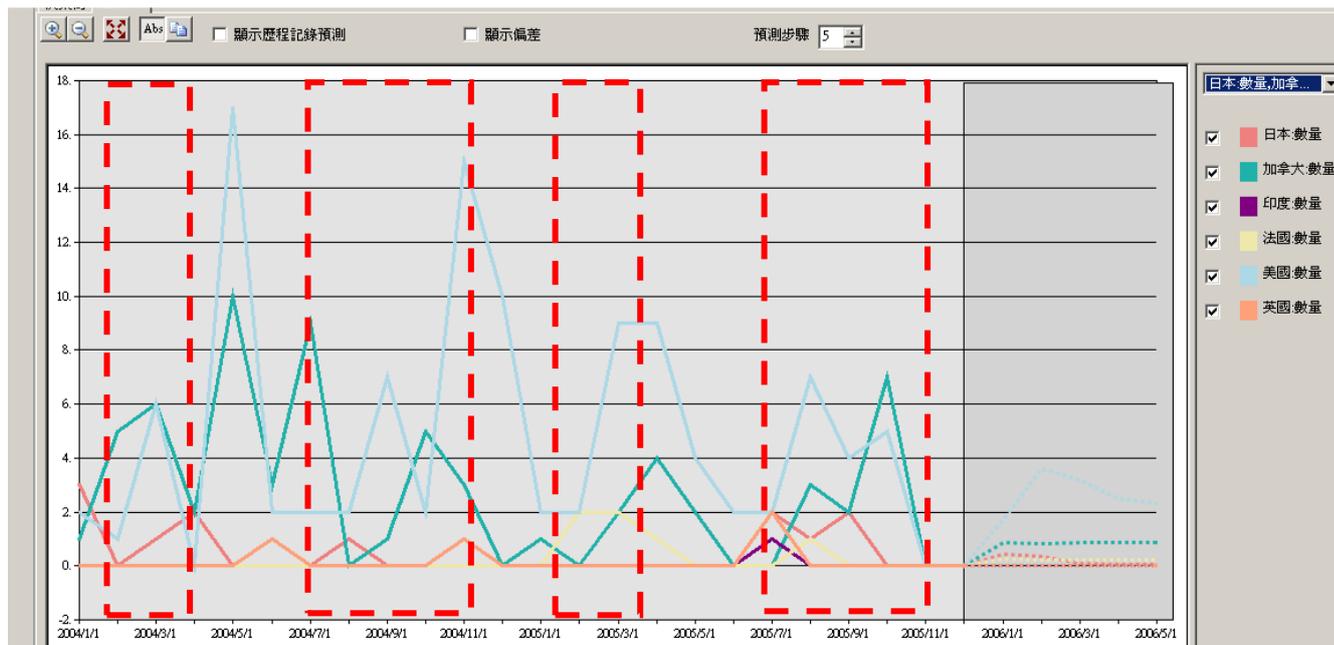


圖23 93/1/1-94/12/31 時間序列分析-1「絕對人數」

資料來源：本研究整理

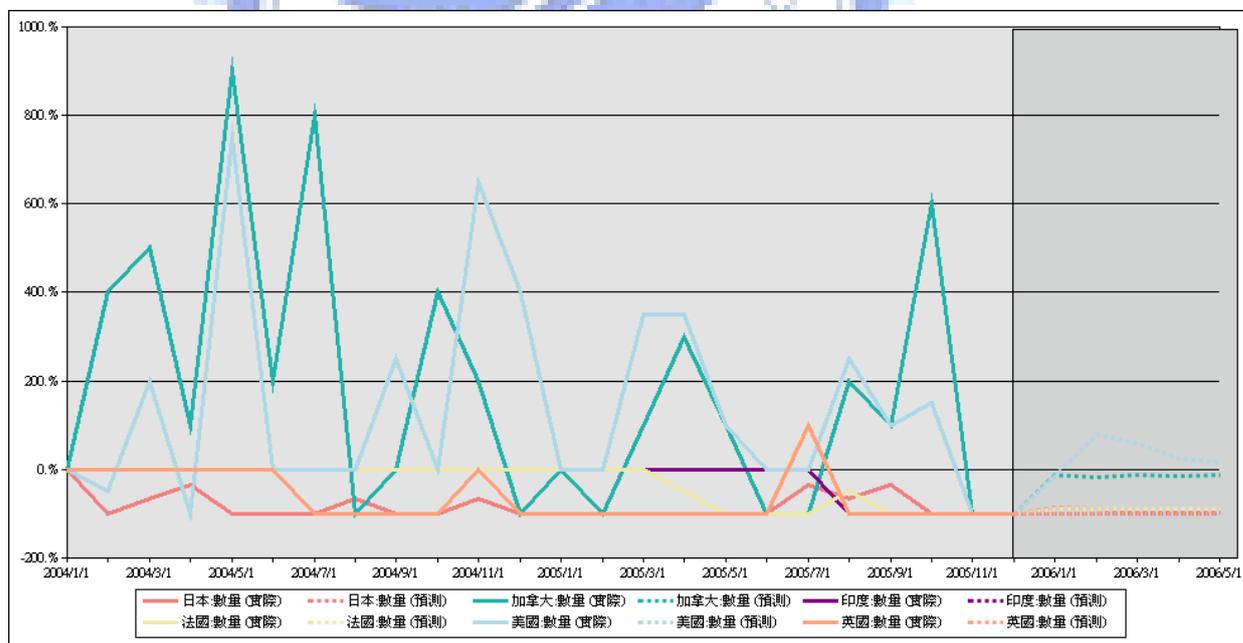


圖24 93/1/1-94/12/31 時間序列分析-2「相對人數」

資料來源：本研究整理

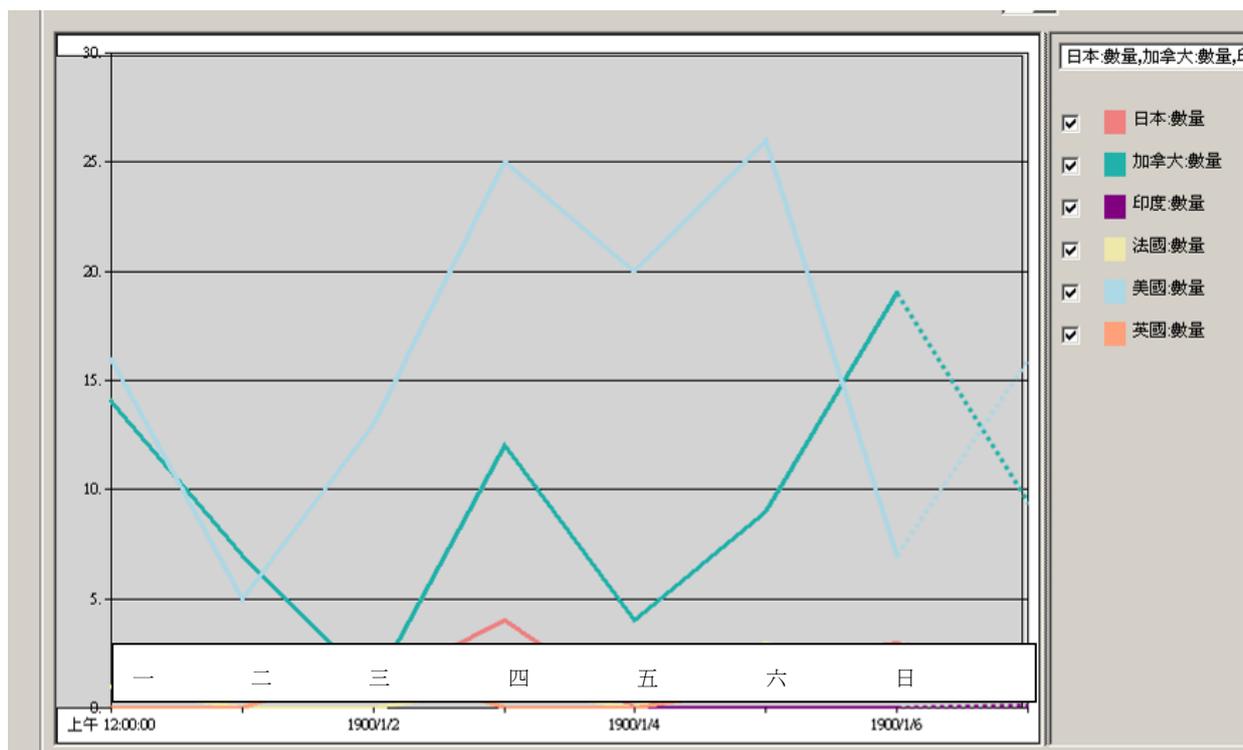


圖25 93/1/1-94/12/3 每週偷渡人數分析

資料來源：本研究整理

另外，針對各時段的偷渡人數分析部分，我們僅針對美、加、日三國進行處理。我們將時間序列設定為 24 小時制，參數 Periodicity_Hint 設定為 {24}。從圖 26 可以看出偷渡人數不分偷渡目的地，尖峰集中在午間 12-15 時及夜間 21 點過後，此固然與客機的預定時刻表極相關，以美國偷渡時段人數曲線為例，飛美航班時刻表 14-16 時共計 3 班，18-19 時共計 10 班，22-24 時為 11 班，而但我們仍應考量搭配偷渡客之抵台、離台航班的銜接來作判斷分析。

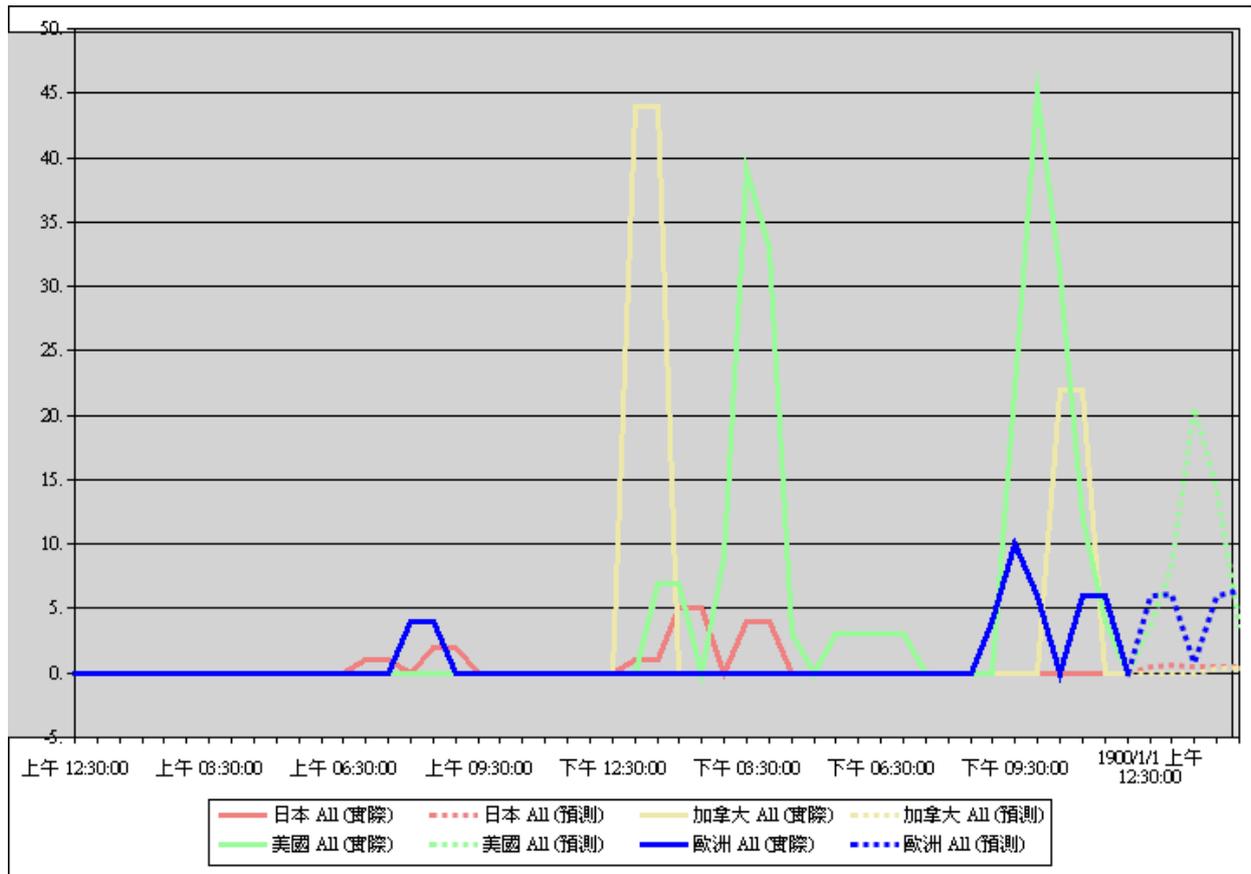


圖26 美、加、日三國 93/1/1-94/12/31 偷渡時段分析

資料來源：本研究整理

4.4.5.2 經濟數據與失業率對於偷渡人數之影響預測

人蛇集團欲經營某一條偷渡路線，必有其市場經營上之考量，我們從前述的分析得知得知偷渡目的地的選擇，有語言、文化上的背景，但更重要的是經濟活動。我們從資料的彙整中得知在台灣機場所查獲的偷渡案件中，以偷渡美、加、日三國的人數最多。我們想瞭解偷渡人數的增減與偷渡目的地國家的經濟是否息息相關。故欲從此角度切入分析。偷渡客既然多以經濟為偷渡的目的，我們假設國內生產毛額、國內總儲蓄額這兩類資料可充分反應我們所要的分析結果。國內生產毛額（Gross Domestic Product 簡稱 GDP）指標：可以瞭解一國境內從事各種經濟活動的總成果，國內生產毛額的大小，代表一個國家的經濟規模；另外，國內總儲蓄額（Gross domestic savings per capit 簡稱 GDS），為該國民眾平

均的儲蓄總額。(年度值，單位：美元)。

上述資料由瑞士洛桑管理學院 (IMD, The International Institute for Management Development) 線上資料庫取得 (以下簡稱: IMD), 而輸入的期間為偷渡資料前一年數據, 以本研究資料為例, 偷渡資料為 2004/1/1-2005/12/31, 而經濟數據則前一年度之 2003/1/1-2004/12/31, 分析之參數設計如表 15 所示。而我們將相關資料經過正規化, 以使所形成的圖形不致於扭曲過大, 分析結果如圖 27 所示。

表15 經濟數據與失業率對於偷渡人數之影響預測參數設定

資料探勘法	時間序列	
演算方式	自我迴歸樹演算法 (ART)	
相關參數	機率>0.8	最小支持度 3
變數名稱	變數型態	資料類型
時間區塊	Key Time	Date
美日加國家(GDP)	Predict	Long
美日加國家(GDS)	Predict	Long
偷渡人數	Predict	Long

資料來源：本研究整理

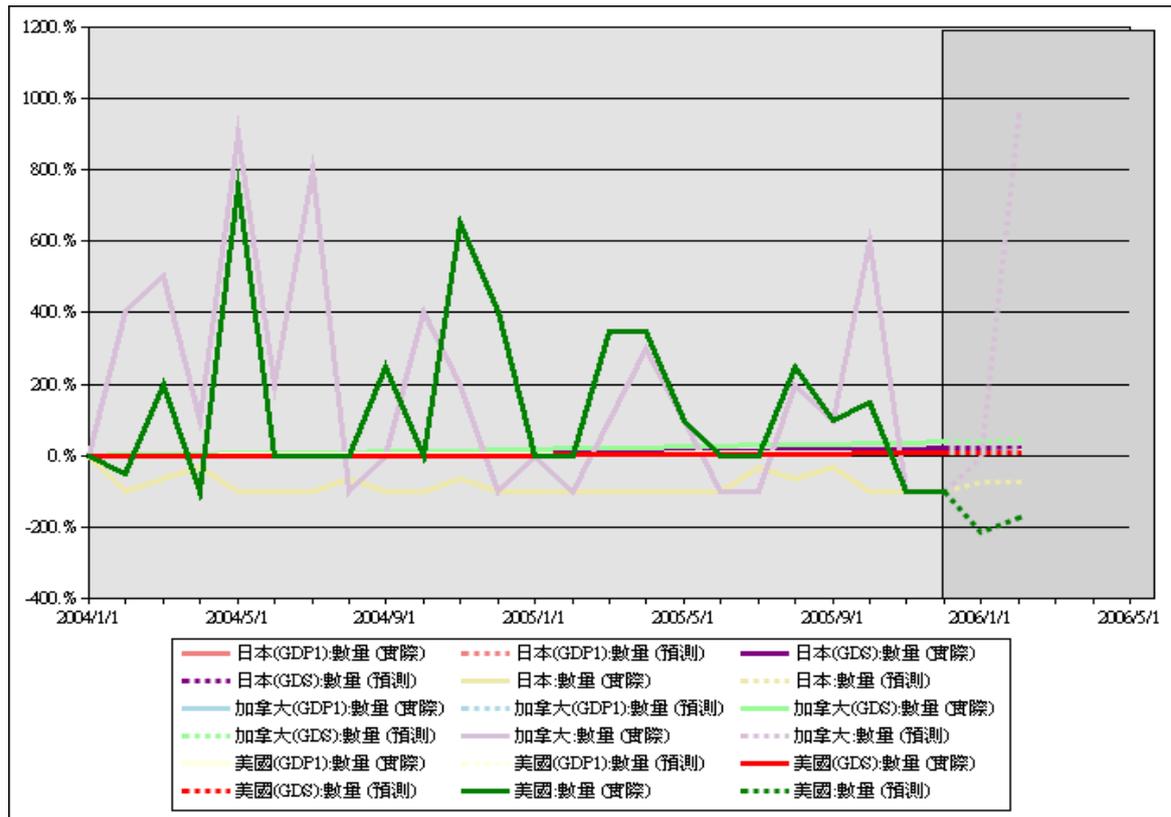


圖 27 美、加、日三國 93/1/1-94/12/3 偷渡人數與經濟數據之影響分析

資料來源：本研究整理

而偷渡過去的目的地國家，偷渡客也許也想知道是不是容易找到工作，故就業是否便利，也成為偷渡客關心的焦點。以下我們從 IMD 線上資料庫取得下列幾個數據 Unemployment rate (失業率；欄位簡稱：UNR)、Long-term unemployment (長期失業率；欄位簡稱：LUN)、Youth unemployment (青年失業率；欄位簡稱：YUN)，上述三個單位皆為百分比。而資料經過正規化，以使所形成的圖形不致於扭曲過大，分析結果如圖 28 所示。從結果我們發現由前一年度的 GDP、GDP、失業率等數據來看，對於判斷偷渡人數的成長情形並沒有太大幫助，僅能瞭解成為偷渡標的之國家的確經濟數據之表現都不錯。而我們認為人蛇集團在安排偷渡航線時，以其在當地所經營之架構為主要考量而招攬客源。部分仍有其國家之歷史、語言、文化、社會福利制度等背景因素，例如：大陸偷渡美國之省籍以福州人居多，其早期偷渡目的以經濟為主，而至今已成為一種當地的時尚虛

榮行為。故本部分研究對於偷渡人數增減的預警部分，因牽涉層面較廣，非從單一經濟角度即可分析出增減趨勢，但從相關經濟數據排名來看，目的地國家的經濟優勢仍是人蛇集團喜好的主要根據。

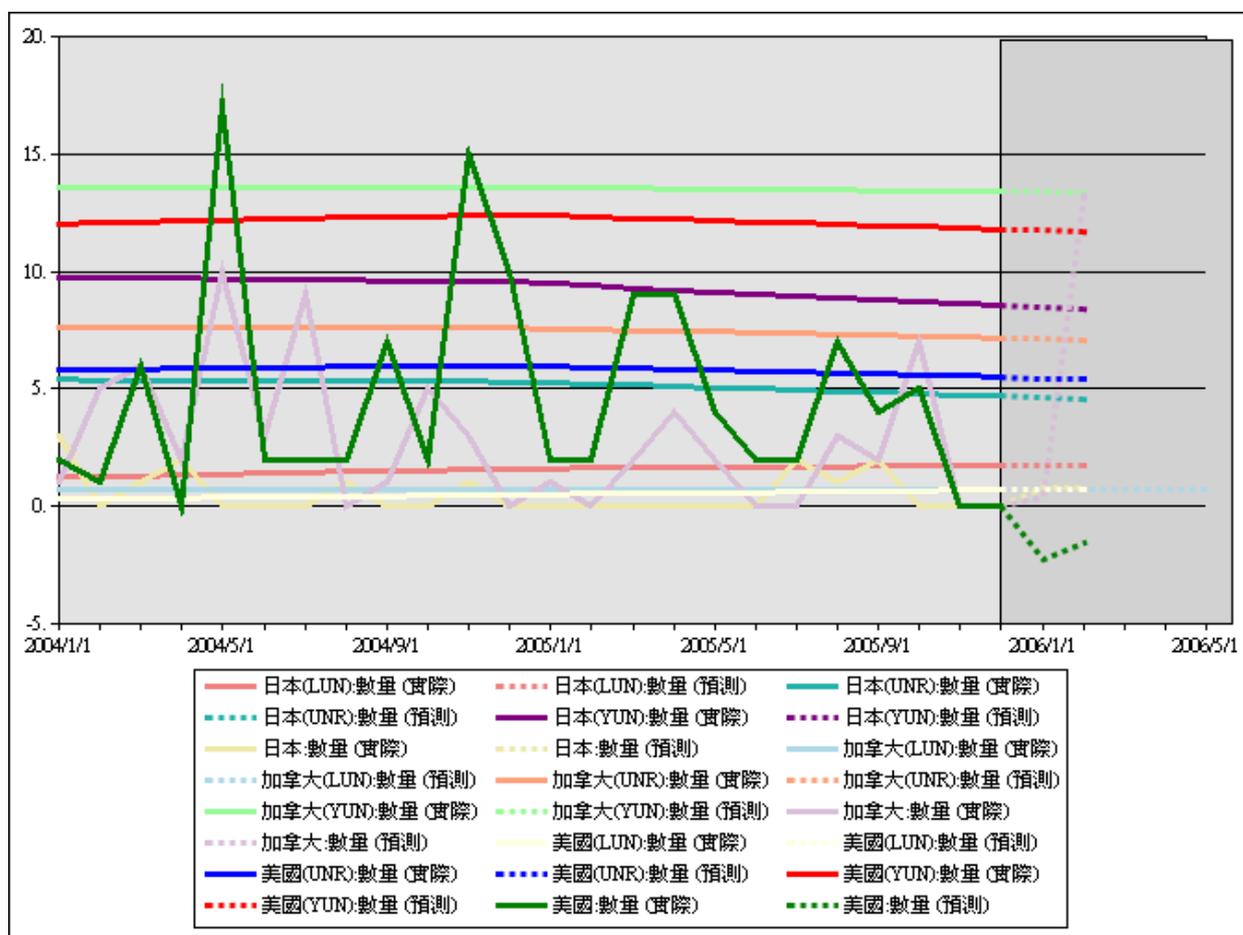


圖28 美、加、日三國 93/1/1-94/12/3 偷渡人數與失業率數據之影響分析

資料來源：本研究整理

五、結論與未來研究方向

5.1 結論

在機場偷渡資料的探勘中，我們使用了三種不同的探勘方式（關聯規則、分群、時間序列），而這三種方式分別是為了能有效協助查緝人員從勤務規劃及執行的角度來打擊機場犯罪偷渡。首先，先針對偷渡行為的特性進行「關聯規則」探勘，由廣而細進行分析。接著，從上述已得到的資訊，由不同面象透過「分群演算法」分析，以獲得各特定群集內所呈現的偷渡活動趨勢。最後，將偷渡活動發生的時間週期及加入經濟及失業率指標，以「時間序列演算法」分析偷渡的行為預測，以完整提供查緝人員於勤務規劃或執行上之參考。

- 1、從關聯規則方法中，我們發現東南亞偷渡客（如：印尼、菲律賓等國）喜以持用自己國籍護照偷渡，且在目的地的選擇上有其特定的喜好。此部分也提醒查緝人員在檢查證件時注意此現象，不可忽略偷渡客來台的原因。
- 2、而以我們設定的 6 個主題進行購物籃關聯分析時，也發現人蛇集團在台偷渡活動之特性。
 - (1)、在「航班號碼」部分：我們發現若一個特定航班，擁有愈多的連結，除了代表多個抵台的航班會往此航班移動，進而讓我們去瞭解航班與航班間之喜好度。其作用可以提供查緝人員擬定查緝策略於特定航班組合的到、離台登機門沿線，尋找於機場管制區遊移的偷渡客。
 - (2)、在「航空公司」部分：我們從分析結果可以得知抵台及離台航班屬於同一個公司或代理公司會增加人蛇集團喜好度，且若分抵台及離台航班屬於同一航廈登機，也相同影響人蛇集團的選擇。

(3)、在「航空代理公司」部分：因台灣偷渡以飛往美、加、日三國的情形較多，而結果也呈現集中在主要負責這些航線的三個公司中。對於這樣的結果，也讓我們對於人蛇集團是否已買通該特定航空代理公司地勤人員之情形持保留態度。

(4)、在「機場」部分：我們發現特定的連結線多的機場，可能已有多組國際性的人蛇集團在當地營運。

(5)、在「國家」部分：我們從分析結果可以得知發現偷渡行為除了以經濟為主要因素外，國家曾被殖民之歷史背景亦為考量之重點。

(6)、在「假照國籍」部分：我們從分析結果可以得知，特定國籍的偷渡客在選擇假證件時，對於同文同種之國籍證件有其優先考量，其次才是該國之經濟背景。

3、而以分群演算法對於偷渡客犯罪手法及選擇目的地機場趨勢，我們得到以下結果：

(1)、在「偷渡犯罪手法趨勢」部分：從分析結果我們得知，群集內以大陸籍偷渡客為主要成員時，因大陸佔地理位置、航班時段之優勢，故進行偷渡較有效率，選擇「海路」、「空路」之搭配也較有彈性，與其他不同群偷渡客所持用的證件種類與證件偽變造手法，皆有明顯相異。此點也可說明偷渡行為可能分屬不同的人蛇集團所安排。

(2)、在「選擇目的地機場趨勢」部分：從分析結果我們得知不同群集內之偷渡客選擇偷渡目的地，有其語言、文化之差異，例如：菲律賓與大陸籍偷渡客就南轅北轍。不過從整體來看，各國偷渡客採行偷渡之手法、路線不同，也難令人察覺出各國人蛇集團在台有跨國性之合作偷渡情形。

4、最後，我們從時序分析來探討偷渡行為發生的頻繁程度，此部分我們用工具所具備的簡潔圖表來呈現各時段、星期、月份偷渡的趨勢，可以發現特定的行為趨勢集中在機場旅客運量高峰期。而對於瞭解偷渡人數之成長情形，搭配經濟及失業率數據，尚不足以完整解釋偷渡人數之成長趨勢，仍需搭配像是偷渡目的國之歷史、語言、文化、社會福利制度等因素，分析效益才會顯著。

5.2 未來研究方向

由於資料探勘在犯罪領域上之應用還不如其他商業來得領域普及，且機場偷渡犯罪行為之紀錄建檔也未普遍化，有很大改善之空間。以下提出幾點建議，作為未來研究之方向：

1、運用文字探勘處理龐大筆錄資料：由於查緝人員移送偷渡犯至司法機關，均需製作筆錄，而筆錄內容除詳細記錄犯罪者之身分資料、犯罪之過程、法律適用等，是極佳之探勘寶庫，而資料分析人員可利用文字辨識軟體加上文字探勘之方式，將過去之筆錄資料進行建檔，轉入已架構好之資料倉儲系統，輔助決策人員制訂勤務政策。

另外，網路上媒體所報導偷渡新聞事件，亦可採用文字探勘方法加上依照事件等級，給予不同的權重，例如：經濟景氣、移民政策等，建立檢索資料庫，以供作為偷渡趨勢之參考。

2、加入犯罪手法細節資料，以輔助查緝決策：在進行查緝機場偷渡犯罪時，考量此類偷渡犯罪有別於一般「海路」、「陸路」偷渡犯罪之特性，為其皆以合法交通工具掩護非法之方式為之，故證件之查驗是佔整個查緝作為中最重要之一環。因從旅客所持有之證件可決定旅客之意圖及身份。

但於台灣機場進行入出境或轉機的旅客，所持有之護照種類眾多且歷年發行之版本種類及防偽功能設計不一，而所採取的偷渡手法也愈來愈多變，加上證件偽、變造手法愈來愈新穎之下，查緝人員很難對所有的手法都能掌握、證件特性都熟

悉，但若能將歷年所查獲之案件犯罪手法，包含其所持用之證件之種類、偽、變造特徵資料等加入資料探勘的分析，將分析出之結果透過專家解析後，將有用之資訊歸納透過 Add-in 的方式，附加於現有的證照查驗系統內，可協助查緝人員以固定式主機或手持式行動設備查驗旅客身份時，輸入對方證件之「國籍」、或「所持有之證件」等條件資料。系統即可從輸入之條件自動判斷可能的犯罪手法，並提示查緝人員應加以注意的部分，可有效減少查緝上漏誤及查驗之時效。此種機制就類似常見的「推薦」機制的作法，只是標的物從購買者的行為變成犯罪者的行為。

- 3、運用社交網絡分析（Social Network Analysis）技術，打擊人蛇集團犯罪：以往查緝各種手法之偷渡犯時，很難將其背後之人蛇集團成員查獲，也導致偷渡犯罪層出不窮，甚至更加劇烈，對打擊偷渡犯罪毫無警示之作用。若能多加蒐集相關數據，使用社交網絡分析（Social Network Analysis）技術，可協助查緝人員識別人蛇集團成員。例如：結合各航空公司航班艙單（此資料會額外記錄來台轉機不入出境之乘客相關資料）、機場通關時間（停留於機場管制區之時間）等資料，以社交網絡分析（Social Network Analysis）進行分析，可有效針對疑似人蛇集團之成員做一過濾之動作，甚至瞭解其於組織內所扮演之角色為何？若經查確認可疑後，可搭配跟監、調閱通聯記錄分析、機場區域監視錄影機資料，有效鎖定人蛇集團成員（含航空公司人員、不肖執法人員等）犯罪之動向，將集團成員一網打盡。

參考文獻

中文部分

- [1] 台灣桃園機場，<http://www.taoyuanairport.gov.tw/web/chinese/schedule/routes.htm>，台灣桃園國際機場-航班資訊-航空公司及航線圖。
- [2] 民航局，<http://www.caa.gov.tw/big5/download/09-01-form20-96.PDF>，台灣地區國際航線班機載客率－中正國際機場。
- [3] 謝立功等，「跨境犯罪偵查之理論與實務」，國科會補助專題研究計畫成果報告，民國 90 年 3 月，第 41-43 頁。
- [4] 何雨畊，「針對向海峽對岸僱兇刑事案件之探討」，民國 95 年。
- [5] 何雨畊，「防制非法入國之探討」，民國 95 年。
- [6] 尹相志，SQL Server 2005 OLAP 線上即時分析，悅知出版社，民國 95 年 11 月。
- [7] 孫惠民，資料採掘理論與實務規劃手冊，松崗出版社，民國 96 年 3 月。
- [8] 謝邦昌，資料採礦與商業智慧-SQL Server 2005，鼎文出版社，民國 94 年 12 月。
- [9] 尹相志，SQL Server 2005 Data Mining 資料採礦與 OFFICE 增益集，悅知出版社，民國 95 年 11 月。
- [10] 尹相志，SQL Server 2005 資料採礦聖經，悅知出版社，民國 95 年 3 月。

[11] 高興宇,「疑有機場地勤人員涉入 台美合逮 5 偷渡陸客」, 聯合報, 民國 96 年 8 月 5 日。

外文部分

[12] Boris Mirkin, Clustering for Data Mining :A Data Recovery Approach, 2005.

[13] Christopher Meek, David, Maxwell Chickering, and David Heckerman, Autoregressive Tree Models for Time-Series Analysis, 2002

[14] Han, J. and M. Kamber, Data Mining : Concepts and Techniques, March 2006.

[15] Tang ,Zhao-Hui and Jamie MacLennan, Data Mining with SQL Server 2005, 2005.

