

資料探勘技術在集集大地震 引致山崩之研究

Discerning Chi-Chi Earthquake-induced Landslide Using Data Mining Technique

鄒明城*

孫志鴻**

Ming-Cheng Tsou

Chin-Hong Sun

Abstract

An earthquake of the magnitude 7.3 on the Richter scale occurred in the middle region of Taiwan on September 21, 1999. This earthquake caused the death of more than two thousand people, severe property loss, and a large number of landslides. A large amount of landslide data and strong motion records were produced by the Chi-Chi earthquake. This research establishes a landslide database and data warehouse collected from the Chi-Chi earthquake in a Geographic Information System (GIS). Using the powerful data-processing function and spatial analysis ability of GIS and Decision Tree Data Mining modeling, we have identified the possible mechanisms that induce landslides and produced a map of landslide susceptibility. Verification proved that the research result is good and the methods used by this research suitable for the forecast of risk assessment of landslide hazard triggered by earthquakes. The findings can be used as a tool for disaster mitigation decision support.

**Keywords: data mining, geographic information system, decision tree,
earthquake-induced landslide.**

* 工業技術研究院能源與資源研究所副研究員

Associate Researcher, Energy and Resources Laboratories, Industrial Technology Research Institute.

** 國立臺灣大學地理環境資源學系教授

Professor, Department of Geography, National Taiwan University.

摘 要

臺灣地處地震帶，部分地區每年均要歷經多次的地震以及周期性的大地震，往往帶來嚴重的山崩以及土石鬆動的現象，若再歷經豪雨後，將帶來嚴重的土石流，造成生命財產的重大損失。1999年9月21日於臺灣中部地區發生芮氏規模7.3的地震，此次強震造成2000多人死亡，難以估計的財產損失，以及數量龐大的坡地災害。而在此次強震中亦獲得大量的坡地破壞資料及地震記錄，可供專家學者進行地震對山崩影響的研究，使得對地震引致山崩的行為有進一步的認識。本研究嘗試以九二一大地震所累積的大批地理資料，經地理資訊系統處理後建立資料倉儲，做為探勘的基礎，再結合地理資訊系統空間資料分析與處理的功能與資料探勘技術中的決策樹 (Decision Tree) 演算法建立預測模式，應用在空間資料庫的知識探索，獲得良好的準確性，期望透過此研究了解引發地震山崩的機制，並且建立山崩災害的潛感圖，提供防災決策支援上的參考。

關鍵字：資料探勘、地理資訊系統、決策樹演算法、地震山崩

前 言

隨著資訊科技的進步，資料的收集紛紛邁入自動化與電腦化，各種資料的收集迅速累積，其數量之大超出過去人類的想像，人類由過去資料貧乏的年代，演進到目前資料與知識爆炸的年代，坐擁如此龐大的資料，其中包含了許多寶貴的知識，如何從這些資料中萃煉出有價值的知識，是當前面臨的一大課題。資料探勘 (Data Mining) 是一種以自動且智慧的方式將資料轉化成有用的資訊和知識 (Chen *et al.*, 1996) 的技術。根據 Frawley 等人 (1991) 的定義：它是從大型資料庫中發現隱含的、先前未知的以及有用的知識。是一個迅速發展的新領域，綜合了機器學習、資料庫、專家系統、模式識別、統計、基於知識的系統 (knowledge-based system) 以及視覺化等領域的相關技術 (Koperski *et al.*, 1996; Miller *et al.*, 2001)，已廣泛的應用在當前許多商業問題的解決上。相同的，在空間問題的研究上也面臨著同樣的問題，遙測影像、GPS 的資料收集、行動通訊裝備以及各種與空間位置相關的紀錄，迅速累積提供我們鉅量的空間資料，它們的處理方式不同於一般的屬性資料，特別著重於空間的意涵，對於空間資料的探勘正逐漸受到重視，研究的方向從基礎空間資料探勘演算法的研究 (Adbelmoty, 1993; Lu *et al.*, 1993; Ester *et al.*, 1997; Koperski *et al.*, 1998; Miller *et al.*, 2000)，到各種空間相關的應用研究，如地圖資訊的擷取 (Malerba *et al.*, 2001)、遙測影像資訊擷取 (Lillesand *et al.*, 2000)、環境特徵的製圖 (Eklund *et al.*, 1998) 時空樣式的擷取 (Openshaw, 1994) 空間互動模式分析 (Arentze *et al.*, 2000; Kwan, 2000; Smyth, 2001) 等，領域十分廣泛。

值此技術不斷獲得空間資訊領域重視之際，1999年9月21日於臺灣中部地區發生芮氏規模7.3的集集大地震，造成2000多人死亡以及難以估計的財產損失和數量龐大的坡地災害。拜現代資訊科技之賜以及國內各單位專家學者積極的投入下，獲得空前未有大量的坡地破壞資料及地震記錄，建立了豐富的地震空間資料庫，提供專家學者進行地震對山崩影響的探討，使得對地震引致山崩的行為有進

一步的認識，也提供本研究進行空間資料探勘上的豐富素材。有鑒於山崩是臺灣最主要的自然災害之一，尤其近年來，不僅人口、產業漸漸由不敷使用之平地逐漸遷移至山坡地，原本就有崩塌、地滑、土石流發生之山坡地，因為人口之進駐，而有生命或財產損失之災害發生。加上頻繁的地震次數以及九二一集集大地震的肆虐，使得原本就脆弱的山坡地地質更容易崩解或分解，以致之後一遇豪雨，土石流災害便不斷發生。為了減少人民生命財產的損失，如何有效的預防地質災害的發生，以達到防災減災的目的已經是相當重要的議題。

在山坡地潛在災害預測的空間問題上，傳統上是由專家學者依經驗以定性分析來判斷，使用現地調查的方式對崩塌因子進行選擇及評分（王鑫，1981；陳振華，1985；張石角，1987；工研院，1998）。此方法的優點為權重給定容易對評分方式不需作其他處理，較為快速，且能夠較切合現地調查結果的重點。但缺點為需要大量的人力和物力，且以人為方式給予權重或評分較為主觀。近年來，隨著地理資訊系統的發展，應用統計學的概念針對各種崩塌因子，給予不同權重之定量分析便逐漸流行。目前，統計學可以說是使用最廣的空間資料分析工具，發展出相當多的演算法與技術，在處理數值資料上已有良好的成效，並且已建立了許多真實空間現象的模式。董啟哲（2001）利用地理資訊系統（GIS）及多變量分析軟體對九二一地震山崩龐大資料進行空間屬性處理及多變量分析。許煜煌（2002）針對九二一地震之山崩以不安定指數法對每個影響因子進行分區與量化的處理，分析各個因子的崩壞比與分區的關係，進而建立彼此的權重關係，最後建立不安定指數法的評估模式。此外，力學模式也另一研究的重點，陳時祖（2000）、陳意璇（2001）針對九二一地震山崩資料，以 Newmark 位移法做為評估邊坡破壞之準確性的依據。Refice 等人（2002）則以 Newmark 位移法結合蒙地卡羅模擬法對義大利南部 Sele 山谷之地震山崩進行研究。而資料探勘及人工智慧技術的興起，則為空間資訊分析開啟了另一扇大門，類神經網路可以說是目前人工智慧應用在空間分析與山崩災害中最常見的一種技術（林祥偉，2001；盧育聘，2002；羅佳明，2002），各類方法都有其適用範圍與相關限制，很難直接比較其優劣。

在面臨本研究區域內龐大的空間資料下，本研究嘗試以可快速建模、預測、提供可了解法則及處理大量資料的資料探勘技術-決策樹（Decision Tree）演算法（Weiss, 1998）做為地震山崩空間資料探勘與分析的工具，並透過地理資訊系統強大的空間資料處理分析能力與關聯式資料庫管理系統，將資料建立成資料倉儲以供做為資料探勘的素材，藉由這樣的技術為資料做精練，自動建立空間知識庫，提供對於地震山崩機制的了解與預測，並建立山崩潛感圖，做為容易發生短週期性大地震或是每年均要歷經多次地震之地區，對於坡地災害防治、環境影響評估以及防災空間決策支援上的參考。

研究方法與地震山崩因子選擇

（一）決策樹演算法

決策樹演算法是一種以遞迴方式把資料集合切分成越來越具同質性次集合的分類演算法，其模式結構以樹狀展開，子節點為切割的資料集，終端節點又稱葉節點，並以能取得最小熵值的節點來做為樹狀結構成長分支的依據（Quilan, 1993），所產生的樹狀結構相較於其他演算法（不論是迴歸類的公式或是類神經網路完全封閉式的黑箱結構），比較容易讓人理解（Openshaw, 1997），這一點其實是相當重

要的，因為在進行研究時，有時候後不只是希望有準確的模型就好，對決策者來說，更重要的是能否從規則的內容中獲得啟發，此時，規則是否能夠以人類所能理解的形式呈現就顯得相當重要了。這也就是為什麼在 Kdnuggets 針對企業資料探勘所做的調查結果顯示，決策樹以高達 16% 的使用比例位居所有資料探勘技術的第一名，而類神經網路則為 9% 的使用率排名第五。

決策樹演算法有相當多的版本，一開始的版本為澳洲學者昆蘭 (Quilan,1986) 所開發之 ID3，之後以此為演化基礎，產生了 C4、C4.5、C5、CART、CHAID、QUEST 等數種修改版的演算法。本研究以 C4.5 (Quilan, 1993) 決策樹演算法來建構預測模式，它採用樹狀分岔的架構來產生規則，適用於所有的分類問題，例如銀行信用卡授信、直銷回應、顧客流失預測、甚至空間議題如本研究之山崩預測議題，皆可應用決策樹演算法來產生容易理解規則。

在分類問題中，最完美的狀況就是將資料分割成一個節點包含了所有 100% 山崩的資料，另一則是 100% 不會山崩的地方，但結果總不會盡如理想，所以不管使用那一個山崩因子來產生子節點，其子節點一定會有分佈不純淨的狀況。因此，我們希望找出那一個山崩因子的效果最好時，就必須判斷根據那一個因子分岔之後，產生的子節點總資料純度最高？此時就產生了幾個分岔 (Split Criteria) 的條件：

1. 純淨度是一個抽象的概念，必須要以數學公式讓他可以量化。
2. 每個因子所產生的分岔數目不同，必須要能夠提供純淨度的加總方式，來評估到底使用哪一個變數能夠讓純淨度最高。
3. 子節點的純淨度總合必須能和母節點的純淨度互相比較，以決定該分岔是否需要保留，因為若是子節點無法顯著比母節點擁有更高的純淨度時，此時該分岔就一點意義都沒有了。
4. 如果輸入變數是連續變數時，如何才能快速搜尋，找出最好的連續變數切點，以達到最高的純淨度。

接著再持續以上的步驟，讓每一個產生的子節點都是做根節點，再繼續試算下一個最佳的分岔因子，繼續產生新的分岔，就這樣，決策樹就會開始生長，一直到子節點再也無法產生純淨度更高的分岔為止。分岔準則是根據增益比值 (gain ratio) 來計算，它首先根據每一節點選擇一個具有最大資訊增益值 (information gain) 屬性，這一組純度尚未淨化的資料集 S 之熵值的計算如下：

$$Entropy(S) = \sum_{i=1}^c -P_i \log_2(P_i)$$

P_i 是指預測因子選項 i 對於分類資料的切分率，即類別 i 在資料集 S 中所佔的比率， c 為類別的數目。

$$Gain(S, A) = Entropy(S) - \sum_v \frac{|S_v|}{|S|} Entropy(S_v)$$

其中， $|S|$ 為資料集 S 樣本數， $|S_v|$ 為資料集中屬性 A 等於 v 的樣本數。

Entropy 表示亂度，這個觀念來自熱力學，用來表示物體分佈的分散狀況，亂度越高，則越無規則，因此決策樹演算法的目標就是希望能夠降低資料分類結果的亂度。比較各個屬性值的增益比值，挑選具有最高增益比值的因子來當做分岔依據。

一但決策樹建立完成後，它可以被轉換成一組組如以下形式的法則：

if

坡度 $< a$

距離車籠埔斷層距離 $< b$

地質分佈 = c

then

山崩預測 = 是

這裏 a、b 為數值資料，c 為類別型資料

但是，如果按照前述流程製作出來的決策樹結構將太過複雜，雖然可以在訓練組產生極佳的效果，但是如果遇到新案例，就會產生過度學習的結果，而使預測能力大大降低。此時就有需要利用修剪技術將模式進行簡化，以設立分類資料切分後之數目的門檻值，將未達門檻之無效或多餘的分岔點修剪掉。因此決策樹演算法可以看作是兩股力量拉扯下的結果，一種是利用因子產生分岔的成長力量，另外是透過修剪來抑制決策樹成長的力量。一棵成長完成的決策樹，除了有完整的樹狀結構可以提供分類規則外，更重要的是他還可以提供機率值。因此我們可以計算葉節點案例的預測變數分佈狀況，例如一條規則中會山崩的機率為 3.5%，而不會山崩為 96.5%，次時這條規則分類的結果是不會山崩，而山崩機率只有 0.035。

由於決策樹演算法需要計算增益比值，有挑選具增益因子的能力，能從一大堆因子中篩選出重要的因子，故可與其它模式進行整合 (Berry, 2000)。本研究所使用的建置方法則是以 Microsoft SQL Server 的 Analysis Service 來建置 Decision Tree 模式，搭配微軟 DSO (Decision Support Object) 之 COM 元件與 VB.Net 程式來操作，其操作語法與 SQL 類似，可方便的進行建模、預測與評估。

(二) 山崩影響因子的選取

回顧以往在大規模的山崩潛感分析研究中，所選用的研究因子一般可分為潛因與誘因，潛因是指坡地本身具有潛在山崩的因子，如岩石的強弱、坡度的陡緩、地層的因素等先天條件，誘因則是直接發生於山崩的環境因素，這些因子在大區域中的分析並不容易考慮，所以大部份是以較容易取得的空間分析因子做為評斷的依據。但由於空間分佈取得的影響因子，絕大部份並非直接代表發生山崩的因素，因此所取得的因子並不是直接反應坡地安全性，而是根據因子本身背後所代表的意義來做間接評估方式 (許煜煌, 2002)。

童啟哲 (2001) 曾利用單因子共變數分析，對相同於本研究範圍之九二一地震山崩資料的高程、坡度、坡向、地質分佈、圖壤無圍壓縮強度、土壤凝距力、土壤摩差角、距離車籠埔層距離、距震央距離、水平地表加速度均方值、垂直地表加速度、水平地表運動方向等十二個因子進行分析發現，高程、坡度、坡向、地質分佈、土壤無圍壓縮強度、水平地表加速度均方值、垂直地表加速度、水平地表運動方向等八個因子具影響力。許煜煌 (2002) 利用單因子顯著性分析，分析相同於本研究範圍之九二一地震山崩資料的坡度、地質分佈、距車籠埔斷層距離、距道路距離、Ia 值、高程、垂直地表加速度、距水系距離、距震央距離、坡向等十個影響因子，發現這十個因子都是引致地震山崩的重要因素。

本研究參考以往學者所採用的影響因子以及資料的取得性，先選取十二個影響因子。另外，考慮到鄰近空間環境的影響性，即山崩往往具有聚集的效果，與週遭環境息息相關，因此帶入空間背景 (spatial context) 的觀念，另外利用其中部份因子再衍生出新的因子，一共是十七個因子進行建模，因子如下：

1. 高程
2. 坡向
3. 坡度
4. 距離車籠埔斷層距離
5. 距離斷層破碎帶距離
6. 距離道路距離
7. 距水系距離
8. 距震央距離
9. 地質分佈狀況
10. 垂直向地表加速度
11. 東西向地表加速度
12. 南北向地表加速度

衍生因子如下：

13. 九格點之平均坡度
14. 九格點之最大最小坡度差
15. 九格點之平均坡向
16. 九格點之最大最小坡向差
17. 地震強度 (Arias Intensity)

以下從地形、地質、區位、地震影響四個方向做探討，分述如下：

(1) 地形因子

一般用以描述坡地之地形者，有坡度、坡向、坡地幾何形狀、坡高等因子，而各地形因子在對地震引致山崩的影響尚有其不同的意義，如：

坡度：坡度的陡峭與否對於山崩具有很大的影響，越陡者造成山崩機率越大。

坡向：坡向本身對坡地影響並非直接，但若考慮坡向與斷層走向的關係，就會對坡地破壞產生影響。

另外考慮到山崩可能會受到周圍環境的影響，例如坡度落差大者易發生崩塌，本研究對周圍環境的狀況 (spatial context) 也加以考慮，以九宮格的方式，求取在 3 × 3 格點範圍內之平均坡度、最大最小坡度差、平均坡向、最大最小坡向差。

(2) 地質因子

一般用以描述該區邊坡本身的材料對抵抗破壞外力的指標，如土壤凝聚力、土壤摩擦角等這些力學性質，在大範圍的研究區內很難取得，而在褶皺、破裂面等地層構造方面，對山崩有一定程度影響，但在大範圍研究上亦不容易取得，本研究以地層性質當作地質特性。

(3) 區位因子

區位因子是指該區受到山崩潛因的影響隨距離變化而有所不同，主要考慮是距離水系遠近對坡地的影響，另外，是否位於斷層破碎帶也是重要的考量因素，若是該處地質條件較差，也較易發生山崩。在考慮到人為因素時，則以距道路距離來代表人為開發對於山崩的影響因子。

(4) 地震影響因子

本研究分別以垂直地表加速度、南北向地表加速度、東西向地表加速度來代表地震時的震動情形，另外以 I_a 值 (Arias Intensity) 來代表崩塌地受到地震能量的影響程度。而在分析地震對山崩的影響上，一般都會將產生嚴重錯動，並且可能是造成此次地震主因的斷層及震央位置納入考慮，同樣地，本研究也將距斷層距離與距震央距離內入考慮。

研究區域與資料的準備

(一) 研究區域

由於九二一大地震所誘發的山崩大多集中在中部區域，故本研究以中部山區為本研究之研究區域。行政區主要包含有臺中市北屯區；臺中縣霧峰鄉、后里鄉、東勢鎮、石岡鄉、豐原市、和平鄉、新社鄉、太平鄉、大里鄉、潭子鄉；南投縣仁愛鄉、國姓鄉、魚池鄉、信義鄉、埔里鄉、草屯鎮、中寮鄉、南投市、集集鎮、水里鄉、名間鄉、竹山鎮、鹿谷鄉等 (如圖 1)。

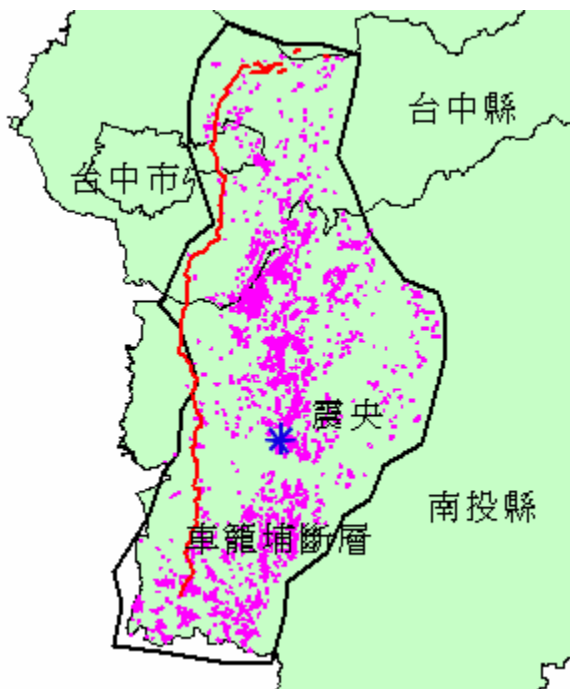


圖 1 研究區域及地震山崩圖

(二) 資料蒐集及處理

本研究所採用的山崩影響因子，包含高程、坡向、坡度、距斷層距離、距斷層破碎帶距離、距震央距離、距道路距離、距水系距離、地層、垂直向地表加速度、東西向地表加速度、南北向地表加速度及 Arias Intensity 等十二個因子及其衍生因子，建立各影響因子所需之基本資料及崩場地資料的收集整理如下；

1. 地震崩場地資料

以工業技術研究院能源與資源研究所（2000）受農委會水土保持局之委託，辦理集集大地震崩場地調查與治理規劃，在經過民國 88 年 4 月 9 日至 7 月 24 日間之災前衛星影像（SPOT）與 9 月 27 日災後之衛星影像及航空照片判釋後，工研院能資所提出了 21969 筆變異點的資料。

2. 數值地形模型

採用由中央大學太空及遙測研究中心所製作的 78 年臺灣地區數值地形模型（DTM），精度為 40m x40m，即每 40m x40m 為一個網格，每個網格會有一個高程值，再經由 ArcView 軟體將高程值轉換成坡度、坡向資料，如此可以獲得高程、坡度、坡向三項因子以及另外衍生出之 9 格平均坡向、9 格最大最小坡向差、9 格平均坡度及 9 格最大最小坡度差。

3. 車籠埔斷層分佈圖

所設定之斷層，是指在集集大地震中產生嚴重錯動，並且是造成這次地震主因的車籠埔斷層，將數化後的斷層圖檔利用 ArcView 軟體產生距離的功能，再產生距斷層距離的圖檔，如此可獲得該項因子的資料。

4. 集集大地震震央位置圖

集集大地震震央位置位於北緯 23.87°，東經 120.75°，轉換為二度分帶座標為 (227591, 2638438)，同樣的，利用 ArcView 產生距離的功能，以震央為中心產生距震央距離圖層。

5. 道路分佈圖

採用內政部所公佈的省縣鄉道路分佈圖數化而成，再以此道路圖利用 ArcView 產生距道路距離圖

6. 活動斷層圖

採用中央大學應用地質研究所繪製之活動斷層圖為基礎，再以 ArcView 產生距斷層破碎帶圖層

7. 水系分佈圖

採用經濟部水利處於民國 88 年所公佈的水細分佈圖數化而成，其水系分佈包含有四級河以上的河川，再利用 ArcView 產生距離的功能，轉換成為距水系距離圖。

8. 數值地質圖

採用經濟部中央地質調查所製作的臺灣地區 1/250,000 數值地質圖為地質分佈狀況的基本資料。

9. 集集大地震地震紀錄

(1) 地表加速度

分別收集了中央氣象局各地震測站的垂直向、東西向、南北向地表加速度資料，並以 ArcView 進行內插分別獲得各方向地表加速度的圖層。

(2) Arias Intensity

引用 Arias (1970) 在對地震儀震動強度進行研究時所提出之地震強度指標，又稱之為 Arias Intensity (Ia)，該指標主要是將在強震延時內的地震紀錄平方之後對時間積分，然後再乘以一個常數而獲得，具有能量的意義。

公式如下：

$$Ia = \frac{\pi}{2g} \int_0^{Td} [a(t)]^2 dt$$

單位: m/sec，其中

a(t)：地震所產生的地表加速度歷時紀錄 (m/sec²)

Td：強震延時 (sec)

g：重力加速度 (m/sec²)

本研究利用程式將各測站東西向及南北向加速度歷時進行合併，計算得到各測站的 Arias Intensity，再以 ArcView 進行內插以獲得 Arias Intensity 圖層。

(三) 空間資料倉儲的建立

在完成以上各圖層之後，再以 ArcView 分別將以上十幾個因子之向量式主題圖層予以網格化，為了配合 DTM 的大小，將每一個圖層的解析度均設為 40m × 40m。之後，再以程式將各個圖層網格結合成一筆一筆的紀錄，每一筆紀錄內的每個欄位均對應到某一個主題圖的格網值，其資料量非常的龐大，共計有約 122 萬餘個網格，其中屬於崩塌的網格大約只有約 6 萬個。然後再透過 SQL Server 之 Data Transformation Service 將資料轉換與清理，以確保資料的適用性，這些紀錄資料在轉進 SQL Server 資料庫管理系統中可成為資料倉儲，透過資料庫管理的功能，可以很方便的提供決策樹模式之建模、驗證與預測之用。

(四) 資料化簡與採樣

雖然，資料探勘是想從大量資料中找尋知識，也就是說資料量越大越有潛力獲得較佳的結果，但是，許多學者 (Weiss, 1998; Berry, 2000; Groth, 2000) 均認為，大量的資料並不能保證獲得比小量資料更佳的預測結果，大多建議以較小量的資料來建立，至少一開始應該如此。除了計算上及時間上的成本太高之外，Weiss (1998) 認為，預測型模式會試著去迎合 (fit) 大量的資料，即便是隨機取樣的資料，也含有許多例外，故資料量越大所含的例外也就越多，為了迎合這樣的資料型態，所得的模式也就越容易出錯。

本研究所面對的資料數高達 122 萬餘筆，其中屬於山崩的部份不到 6 萬筆，為了處理這樣大量的資料，即便是目前的電腦軟硬體技術，處理上亦力有未逮。因此需要以採樣的技術來簡化資料，但是，山崩資料相對於母體而言過於稀少，若採一般隨機抽樣，可能造成對於稀少事件的不易掌握。Berry (2000) 建議可以採用 Oversampling 技術，增加樣本中稀有事件的比率，他認為維持在 10% 至 40% 的比率，通常可以獲得不錯的結果。故本研究將資料分成山崩與未山崩二部份，為了方便於 Excel 中做分析，因此分別從屬於山崩的六萬筆資料中隨機抽樣 2 萬筆、另從未山崩之 116 萬筆資料中抽樣 4 萬

筆混合成 6 萬筆樣本，再將這 6 萬筆資料隨機分成二部份各三萬筆，分別做為建模的訓練 (training) 與測試 (test) 資料。

結果與討論

(一) 模式評估

本研究先以六萬筆採樣資料集中的三萬筆訓練資料集做為模式建構的基礎，待模式建置完成後，再以其它另外三萬筆資料做為模式驗證與評估依據，驗證的方式如下：

1. 錯差矩陣 (confusion matrix)

資料探勘上有所謂的 commission error 及 omission error, commission error 就是把未山崩的資料預測為山崩，將原為山崩資料當作未山崩則是 omission error。對於本研究來說，通常 omission error 的成本或效益要比 commission error 來的高，因此評估本研究模式的第一步，就必須先從錯誤狀態的分類著手，而使用的第一個工具就是錯差矩陣。表 1 為決策樹演算法在三萬筆測試資料集作驗證時所獲得的錯差矩陣，模式的準確率為 86.5%。另一個重要的指標為反查 (Recall)，代表在所有實際山崩中有多少比率為模式正確預測出來，在本研究中反查為 88.8% (表 1 中 $A/(A+B)$)，即 88.8% 的山崩為本模式所正確預測出來。

表 1 錯差矩陣

	實際山崩	實際未崩
預測山崩	12837(A)	2451
預測未崩	1611(B)	13101

2. 累積增益圖 (gain chart)

累積增益圖是一種經常用於資料探勘上的評估圖，它的橫軸及縱軸都是百分比構成的。橫軸百分比代表資料探勘模式機率從高至低排序後的山崩數佔樣本數百分比。縱軸代表回應的山崩數佔總體山崩數的百分比。圖中中間那條 45 度線，代表隨機未做模式的狀態，當我們篩選一半時就剛好有全體一半的效果，曲線越向上彎曲代表模式效果越好，當模式接近 45 度時，代表模式比隨機好不了多少。圖 2 為決策樹演算法的累積增益圖，大約在機率最高的前 40% 預測中就已掌握了 80% 的山崩資料，顯示模式的預測效果不錯。

(二) 空間背景 (Spatial Context) 資料的加入對於模式的影響

加入 spatial context 的計算後，對於模式具有提高預測正確率的效果，可從 82.3% 提升至 86.5%。

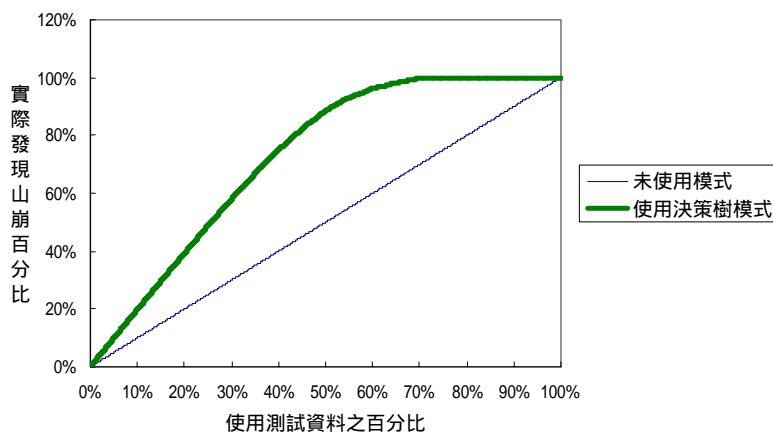


圖 2 決策樹模式累積增益圖

(三) 山崩影響因子之分析

透過決策樹演算法對於各個因子的增益值之計算，可以找出各個因子對於山崩影響性的大小。本研究的做法是以 SQL Server 之 Analysis Service 中決策樹演算法的網路依存分析用視覺化的方式呈現出來。分析顯示，除了與坡向有關的因子外，其餘的輸入因子均對地震山崩均有所影響，其影響程度各不相同，程度由大至小分別為地質分佈、九格平均坡度、距車籠埔斷層距離、Ia 值、南北向地表加速度、東西向地表加速度、垂直地表加速度、距震央距離、高程、九格最大最小坡度差、距道路距離、距水系距離、距斷層破碎帶距離、坡度。

這樣的結果與許煜煌 (2002) 利用不安定指數法分析所獲得的結果類似，在他的研究中，因子影響呈度的排列為坡度、地質分佈、距車籠埔斷層距離、距道路距離、Ia 值、高程、垂直地表加速度、距水系距離、距震央距離、坡向。比較二者的研究均顯示，地質、坡度與距車籠埔斷層距離為影響地震山崩的最重要因子，但是在坡度的分析上，本研究之推論為九格平均坡度較單一格點坡度有更高的影響力，將週遭坡度一併計算對於面積較大的崩塌地比較合理。二者均將坡向視為影響最小的因子，本研究甚至把坡向因子排除。另外，本研究在地表加速度的分析上，三類地表加速度有類似的影響程度，或許可以簡化僅以某一值代表。比較二者的正確率來看，本研究有 86% 的正確率，相較於不安定指數法 68% 的正確率，因此本研究具有相當高的參考價值。

(四) 法則的產生

由於決策樹會產生相當多的法則構成一個知識庫，在此不一一列舉，僅列出部份法則供參考，其形式如下：

例如

if (13.8 公里 < 距車籠埔斷層距離 <= 16.4 公里或距車籠埔斷層距離 > 17.7 公里) and (地質分佈 = 頭料山層及卑南山礫岩) and (324 gal < 垂直地表加速度) and (九格平均坡度 >= 25)

then

山崩預測 = 是 (信賴度 98.7%)

(五) 山崩潛感圖的製作

決策樹演算法雖然可以提供較為直覺的法則，但是當法則太多或太長時間不易於決策者的決策判斷。本研究透過決策樹演算法對於研究區內所有格點預測其山崩機率值，然後再將這些值輸入 ArcView 中，並且依值的大小分為五級（1 至 0.8 為高，0.8 至 0.6 為中高，0.6 至 0.4 中，0.4 至 0.2 中低、0.2 至 0 為低），以山崩潛感圖來呈現，圖 3 與圖 4 分別為研究區的山崩潛感圖，可以看出模式根據背景條件所算出的機率，大部份的山崩都分佈在高與中高潛感區，只有極少數山崩分佈在中的地方，合理的推論出各個山崩潛感區。更有價值的是，在原始向量資料網格化的過程中，部份面積過小的山崩區往往會被排除掉，但透過機率圖的展現後，這些山崩點又可以合理的表示出來，故發現以機率圖來呈現較非有即無二分式預測圖更為客觀 (Refice *et al.*, 2002)，可應用於山坡整治或防災規劃上。

結論與建議

(一) 面對如本研究區這樣的大的研究範圍，適當的抽樣是必須的。本研究對於相對稀少的山崩資料以 Oversampling 方式進行採樣，再透過資料探勘技術中的決策樹演算法進行預測，於地震山崩預測模式可以獲得 86.5% 正確率。

(二) 在加入 spatial context 因子的考慮後，對於正確率有明顯的提升，顯示空間鄰近關係的確有所影響，本研究僅先考慮影響較大的九格坡度、坡向之最大最小值與平均值等地形因子，未來可進行其他因子在鄰近影響上的分析。

(三) 透過決策樹演算法的增益比計算後發現，除了與坡向有關的因子外，其餘的輸入因子均對地震山崩有所影響，但其影響程度由大至小分別為地質分佈、九格平均坡度、距車籠埔斷層距離、Ia 值、南北向地表加速度、東西向地表加速度、垂直地表加速度、距震央距離、高程、九格最大最小坡度差、距道路距離、距水系距離、距斷層破碎帶距離、坡度。

(四) 本研究透過預測的機率轉化為山崩潛感圖，獲得較預測圖合理的表現方式，這樣的方法論對於經常發生地震或有低迴歸期大地震的地區，具有山坡地災害防治決策上的參考。未來的研究可以考慮將防治成本與效益等因子建置於地理資訊系統中，透過資料探勘中利潤分析的技術，排定符合工程經濟上優先整治的依據。

(五) 由於本研究是以九二一大地震區域相關資料進行歸納而建立模式，故提出的模式適用範圍應為研究區域附近，未來遭受類似斷層錯動引起地震影響下，所引致山崩危險的參考。且九二一地震係由車籠埔斷層錯動所引起的，本研究所建立的預測模式尚無法判定是否適用於非斷層錯動所引致的地震。但相同的方法論仍可應用於其他地區預測模式的建立上。

(六) 崩塌有其方向性，造成某處的崩塌，其弱面也許僅是坡腳之局部區位。因本研究係以網格式資料做為建模的基礎，崩塌的計算範圍擴及至整個坡面，在方向性的探討上有所限制，未來可結合向量式資料，將空間物件間的位相關係納入考慮。

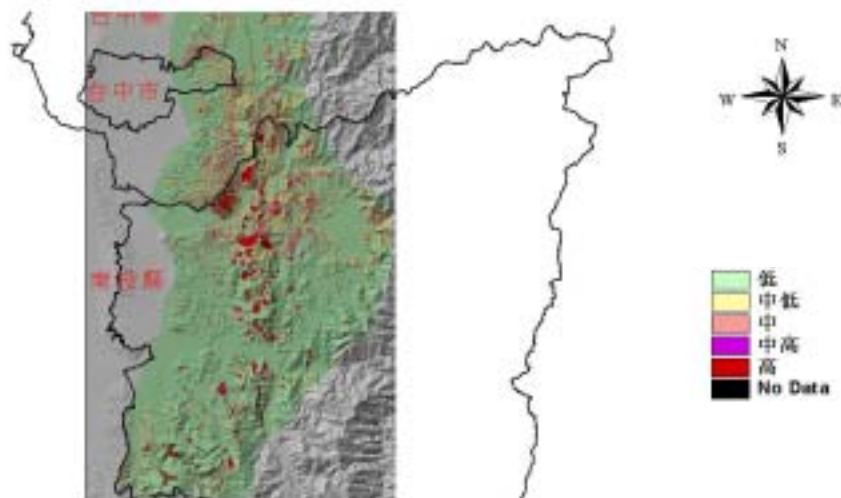


圖 3 全區山崩潛感圖

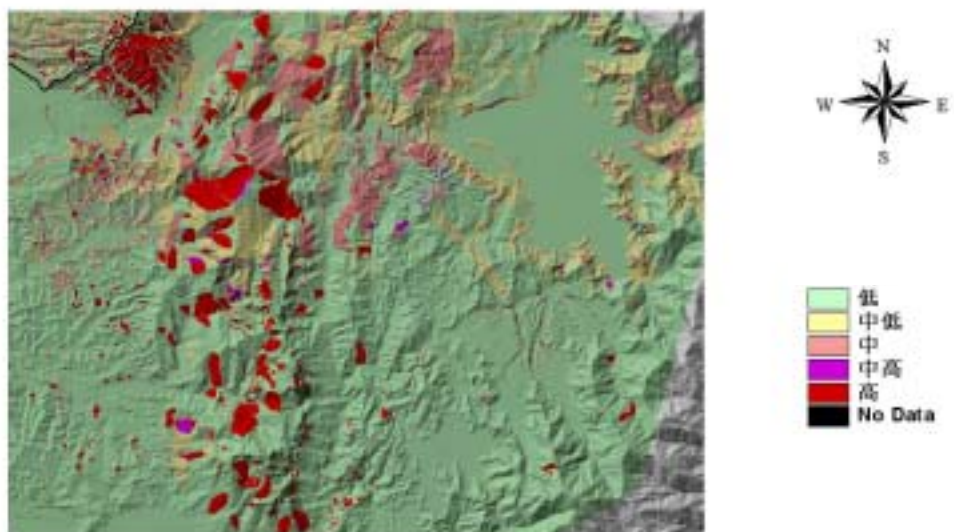


圖 4 山崩前感圖放大

(七) 山崩事件發生的機率有時會與規模成反比，即地質破碎不穩定的地區，因平時較常發生山崩，故山崩的規模有可能反而較小。本研究礙於大範圍資料收集上的困難，所使用的因子仍屬有限，未來若能加入如植被、覆土厚度、覆土性質、地下水狀況等因素，當能提供更具說服力的結果說明。

引用文獻

- 工業技術研究院能源與資源研究所 (1998) 臺灣省重要都會區環境地質資料庫，能資所報告編號：063-88-L029。
- 張石角 (1987) 山坡地潛在危險之預測及其在環境影響評估之應用，*中華水土保持學報*, 18(2): 41-62。
- 童啟哲 (2001) 應用地理資訊系統於地震引致坡地破壞多變量模式分析，國立臺灣大學土木工程研究所碩士論文。
- 許煜煌 (2002) 以不安定指數法進行地震引致坡地破壞模式分析，國立臺灣大學土木工程研究所碩士論文。
- 陳振華 (1984) 衛星與雷達影像之工程地質判釋，遙感探測，8: 86-96。
- 王鑫 (1986) 中橫公路道路邊坡的地貌分析，行政院國家科學委員會防災科技研究報告 74-48 號，72。
- 陳時祖 (1996) 賀伯颱風造成山區道路嚴重受損之成因及減輕未來災害之探討 - 以新中橫公路為例，國科會工程處工程科技推展中心，賀伯颱風災害調查研會論文集，1-10。
- 陳意璇 (2002) 溪頭地區山崩潛感圖製作，國立臺灣大學土木工程研究所碩士論文。
- 林祥偉 (2003) 地理資訊系統與人工智慧之整合研究，國立臺灣大學地理環境資源研究所博士論文。
- 盧育聘 (2003) 類神經網路於公路邊坡破壞潛能之評估，國立中山大學資源環境研究所碩士論文。
- 羅佳明 (2003) GPS/GIS/RS 應用於地震災區坡地災害防治工程調查及其風險評估模式之建置與應用，屏東科技大學土木工程研究所碩士論文。
- 工業技術研究院能源與資源研究所 (2000) 九二一震災系列調查 (一) 崩塌地調查治理規劃，行政院農業委員會水土保持局。
- 尹相志 (2003) *SQL 2000 Analysis Service 資料採礦服務*，台北：維科圖書公司，313。
- Adbelmoty, A. I., Williams, M. H. and Paton, N. W. (1993) Deduction and deductive databases for geographic data handling, *Advance in Spatial Databases. Lecture Notes in Computer Science*, 692, Berlin: Springer-Verlag, 441-464.
- Arentze, T. A., Hofman, F., van Mourik, H., Timmermans, H. J. P. and Wets, G. (2000) Using decision tree induction systems for modeling space-time behavior, *Geographical Analysis* 32(4) : 52-72.
- Arias, A. (1970) A measure of earthquake intensity. In: Hansen, R. J. (ed.) *Seismic Design for Nuclear Power Plants*, Cambridge, Massachusetts: MIT Press, 438-483.
- Berry, M. J. A. and Linoff, G. S. (2000) *Mastering Data Mining*, New York: Wiley.
- Chen, M. S. Han, J. and Yu, P. S. (1996) Data mining: an overview from a database perspective, *IEEE Transactions on Knowledge and Data Engineering*, 8(6) : 632-650
- Ester, M., Kriegel, Hans-Peter and Sander, J. (1997) Spatial data mining: a database approach, *Proc. 5th*

- International Symp. On Spatial Database (SSD'97)*, 47-66.
- Eklund, P. W., Kirkby, S. D. and Salim, A. (1998) Data Mining and soil salinity analysis, *International Journal of Geographical Information Science*, 12: 247-68.
- Frawley, W. and Piatetsky-Shapiro, G. (1991) Matheus, C. Knowledge discovery in database: an overview, *Knowledge Discovery in Database*, Cambridge, Massachuseffs: AAAI/MIT Press.
- Groth, R. (2000) *Data Mining-Building Competitive Advantage*, New York : Prentice Hall.
- Koperski, K., Adhichary, J. and Han, J. (1996) Spatial data mining: progress and challenges survey paper, *SIGMOD'96 Workshop on Ressearch Issues on Data Mining and Knowledge Discovery*.
- Koperski, K., Adhichary, J. and Han, J. (1998) Mining knowledge in geographical data, *Communication of ACM*, <http://db.cs.sfu.ca/sections/publication/kdd/kdd.html>.
- Kwan, M. P. (2000) Interactive geovisualization of activity-travel patterns using three dimensional geographical information systems: a methodological exploration with a large data set, *Transaction Research C*, in press.
- Lillesand, T. M. and Kiefer, R. W. (2000) *Remote Sensing and Image Interpretation*, 4ed, New York : John Wiley.
- Lu, W., Han, J. and Ooi, B. C. (1993) Discovery of general knowledge in large spatial database, *Proc. Far East Workshop on Geographic Information Systems*, 275-289.
- Malerba, D., Esposito, F., Lanza, A. and Lisi, F. A. (2001) Machine learning for information extraction from topographic maps, *Geographical Data Mining and Knowledge Discovery*, 291-314.
- Miller, H. J. and Han, J. (2001) *Geographical Data Mining and Knowledge Discovery*, Lodon: Taylor and Francis, 3-31.
- Openshaw, S. (1994) Two exploratory space-time-attribute pattern analyzer relevant to GIS, *Spatial Analysis ans GIS*, London: Taylor and Francis, 83-104.
- Openshaw, S. and Openshaw, C. (1997) *Artificial Intelligence in Geography*, New York : John Wiley and Sons.
- Quinlan, J. R. (1986) Induction of decision tree, *Machine Learning*, 1: 81-106.
- Quinlan, J. R. (1993) *C4.5: Programs for Machine Learning*, New York : Morgan Kaufmann.
- Refice, A. (2002) Probabilistic modeling of uncentainties in earthquake-infuced landslide hazard assesment, *Computer and Geoscience*, 28: 735-749.
- Smyth, C. S. (2001) Mining mobile trajectories, *Geographical Data Mining and Knowledge Discovery*, London: Taylor and Francis, 337-361.
- Weiss, S. M. and Indurkhya, N. (1998) *Predictive Data Mining-A Pratical Guide*, New York: Morgan Kaufmann.

92年12月17日 收稿

93年3月9日 修正

93年4月25日 接受