
**巨量資料應用於坡地智慧防災系統建構
之先驅研究-以土石流災害為例**

**A pioneering study of the application of big
data on the establishment of smart system
for slopeland disaster prevention**

執行 單位：國立臺灣大學

執行 期間：105 年 04 月 01 日至 105 年 12 月 31 日

計畫主持人：國立臺灣大學生物環境系統工程學系 范正成教授

行政院農業委員會水土保持局 編印

中華民國 105 年 12 月

（本報告書內容及建議純屬執行單位意見，僅供本局施政參考）

摘 要

台灣地區自民國 91 年迄今，水土保持局陸續建置完成全台灣 24 個固定式土石流觀測站，並設置各式監測儀器。本研究以此為基礎，透過巨量資料之概念，將所有土石流觀測站之大量監測資訊納入，進行與土石流發生關聯性之分析，建構土石流災害之智慧防災系統。在本先驅研究中，嘗試利用時序性資料的特徵學習法(即詞袋模型，Bag of Word)，將巨量的監測資料重新解讀，並可將資料進行降維，以提升模式運算效率，進而搭配異常事件偵測演算法，進行土石流警戒模式之建置。透過陳有蘭溪流域四個土石流觀測站的實際測試結果可知，本研究在土石流發生警戒之命中率與可信度的表現，均較現行土石流預警方式為高。未來可以此模式為基礎進行發展，藉以提升災害預測的準確性，達到健全坡地災害防災應變機制的最終目的。

關鍵詞：土石流觀測站、巨量資料、詞袋模型、土石流警戒

Abstract

Since 2002, the Bureau of Soil and Water Conservation has installed 24 fixed type stations for debris flow observation as well as various monitoring devices in Taiwan. Applying the concept of big data, all the data obtained from the monitoring systems of the debris flow observation stations were used to analyze the relationships among the data and debris flow occurrence to establish the smart system for slopeland disaster prevention. In this study, a method of sequential feature learning algorithm (ie. a model of Bag of Word) was tried to re-interpret monitored big data, followed by dimension reduction to enhance computational efficiency, and accompanied with abnormal event detection algorithm, to establish a debris flow warning model. From the actual test results obtained from 4 debris flow observation stations in the watershed of Chenyulan stream, the hitting rate and credibility of debris flow warning proposed in this study are higher than that currently used. It is suggested that the proposed model might be applied to increase the accuracy of disaster prediction, and to build a sound response mechanism for slopeland disaster prevention.

Keywords: debris flow observation station, big data, bag of word, debris flow warning

目次

第一章 前言	1
第一節 計畫緣起及目的	1
第二節 研究範圍	2
第二章 研究方法與步驟	3
第一節 土石流觀測站之各式監測儀器說明	3
第二節 異常事件偵測系統之發展	6
第三節 異常偵測模型之架構與設定	14
第四節 模式評估	15
第三章 結果與討論	18
第一節 不同監測因子之測試探討	18
第二節 異常偵測模型之初步測試	21
第三節 土石流預警系統模式建置	23
第四節 土石流預警系統模式測試結果與討論	25
第四章 結論與建議	34
參考文獻	35

表次

表 2-1 上安土石流觀測站之基本資訊.....	4
表 2-2 郡坑土石流觀測站之基本資訊.....	4
表 2-3 神木村土石流觀測站之基本資訊.....	5
表 2-4 豐丘土石流觀測站之基本資訊.....	5
表 2-5 土石流觀測站之監測儀器說明.....	6
表 3-1 神木村觀測站之模型測試結果一覽.....	21
表 3-2 各測站經試驗後之最佳參數設定.....	26
表 3-3 命中率與可信度分析.....	26
表 3-4 神木站訓練時間長度之結果比較.....	28
表 3-5 神木站試驗最佳參數設定與土石流事件及各警報比較.....	28
表 3-6 豐丘站訓練時間長度之結果比較.....	30
表 3-7 豐丘站試驗最佳參數設定與土石流事件及各警報比較.....	30
表 3-8 上安站訓練時間長度之結果比較.....	31
表 3-9 上安站試驗最佳參數設定土石流事件與各警報比較.....	31
表 3-10 郡坑站訓練時間長度之結果比較.....	32
表 3-11 郡坑站試驗最佳參數設定土石流事件與各警報比較.....	32

圖次

圖 1-1 陳有蘭溪流域測站分佈圖.....	2
圖 2-1 計畫架構流程.....	3
圖 2-2 加入或者移除異常點對主成分方向的影響.....	7
圖 2-3 利用複製異常點加大對主成分方向的影響.....	8
圖 2-4 傳統之時間序列向量化之方式.....	10
圖 2-5 生成字典之示意圖.....	10
圖 2-6 利用字典來對時間序列作編碼之示意圖.....	11
圖 2-7 利用序列資料特性作向量化之資料訓練與預測架構圖.....	12
圖 2-8 土石流發生之異常偵測模型架構.....	14
圖 2-9 異常偵測模型之警戒發布與解除架構.....	15
圖 2-10 水保局有效累積雨量及紅色警戒發布模式.....	16
圖 3-1 土壤含水量及雨量關係之示意圖.....	19
圖 3-2 地聲訊號與土石流發生時間關係之示意圖.....	20
圖 3-3 流速與土石流發生之關係示意圖.....	20
圖 3-4 不同組別偵測之異常行為特徵.....	22
圖 3-5 特徵學習法模型之漏報事件.....	23
圖 3-6 模式操作流程.....	24
圖 3-7 詞袋模型漏報之土石流事件.....	29

第一章 前言

第一節 計畫緣起及目的

傳統上，土石流發生之警戒模式大致可分為事前預警及事發偵測等兩種。其中，事前預警係利用觸發土石流發生之因子進行預測，基於雨量資料取得的方便性，一般而言皆以「降雨量」及「降雨延時」作為土石流發生之預警指標，如 Caine(1980)、Cannon & Ellen(1985)、Wieczorek et al. (1987) 及 Keefer et al. (1987)等，皆在不考慮前期降雨(The antecedent precipitation)的情況下，利用激發土石流該降雨事件之降雨強度及降雨延時，建立土石流發生之臨界降雨線；而范正成等(1999、2001)以洪水消退係數(recession coefficient)在考量臨前降雨之影響下，探討有效累積降雨量與降雨延時之關係。詹錢登(2003) 利用全台各地之土石流發生事件及降雨資料為基礎，以統計方式建立臺灣機率式土石流發生降雨警戒模式及土石流發生降雨警戒基準值，此亦為現今台灣地區進行土石流警戒發佈之依據。

此外，事發偵測方面，則是利用鋼索檢知器、地聲、土壤含水量或影像判識等儀器，依據土石流發生的特性，達到觸發警報，甚至提前預警之目的。目前世界各國於土石流事發偵測之方式，皆以土石流自動化觀測系統為主。在台灣地區，自民國 91 年迄今，水土保持局陸續建置完成全台灣 24 個固定式土石流觀測站，其監測儀器包括雨量計、地聲感測器、土壤含水量計、鋼索檢知器、流速計、水位計及 CCD 影像監視器等，雖許多相關研究已針對上述儀器偵測之結果與土石流發生間的關係進行探討(如林炳森，1999、張守陽等，2000、劉格非，2000 及黃清哲等，2004 等)，然而該些研究大多僅以單一指標來進行與土石流發生關聯性的探討。同時，這些研究雖皆具有一定程度之成果，但目前大多仍未達到實際應用之層面，顯然在預測準確性或實際預警操作層面，仍有持續精進且改善之空間。

緣上述，本研究透過巨量資料之概念，將所有土石流觀測站之大量監測資訊均納入分析。透過訊號異常事件偵測系統之開發，進行與土石流發生關聯性之分析。值得一提的是，在巨量資料分析的概念中，先不細究因子與土石流發生現象之因果關係或物理意義，而探討因子與現象之相關性，希冀發展出有別以往且全新概念的土石流警戒系統。其主要目的係為結合土石流觀測站的大量監測資訊之分析應用，嘗試建構土石流災害之智慧防災系統，藉以提升災害預測的準確性，達到健全坡地災害防災應變機制的最終目的。

第二節 研究範圍

本案研究範圍為陳有蘭溪流域，以上安、郡坑、豐丘及神木村等四個土石流自動化觀測站之監測資料，分別進行土石流預警系統的試驗。流域內之測站分佈如圖 1-1 所示。

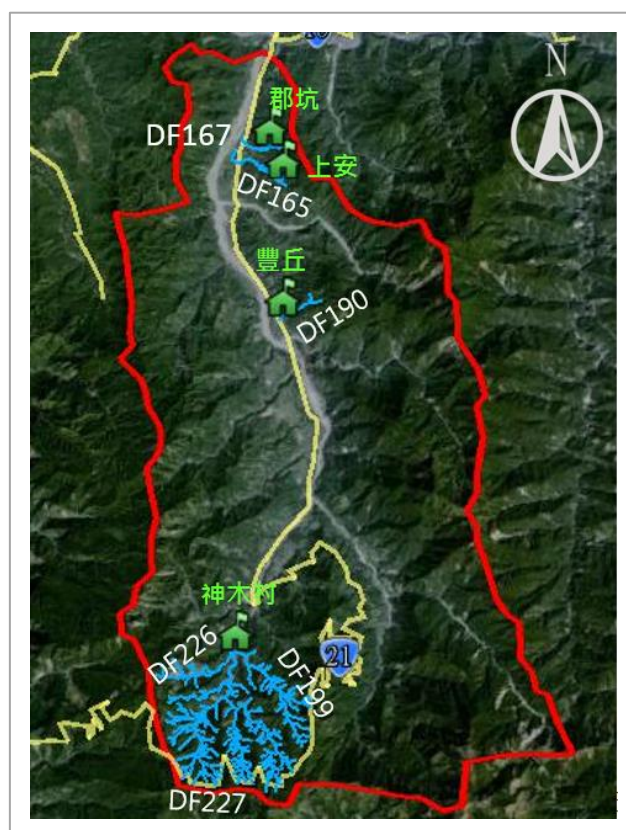


圖 1-1 陳有蘭溪流域測站分佈圖

第二章 研究方法與步驟

為從大量監測資料分析出土石流發生之相關因子，並且運用該因子發展土石流警戒系統，本研究主要之工作包括：(1)土石流觀測站之監測資訊蒐集、(2)監測資料過濾與標記、(3)異常事件偵測演算法之開發及(4)異常事件偵測系統之測試與驗證等，整體研究架構如下圖所示。細部工作內容則說明如後。

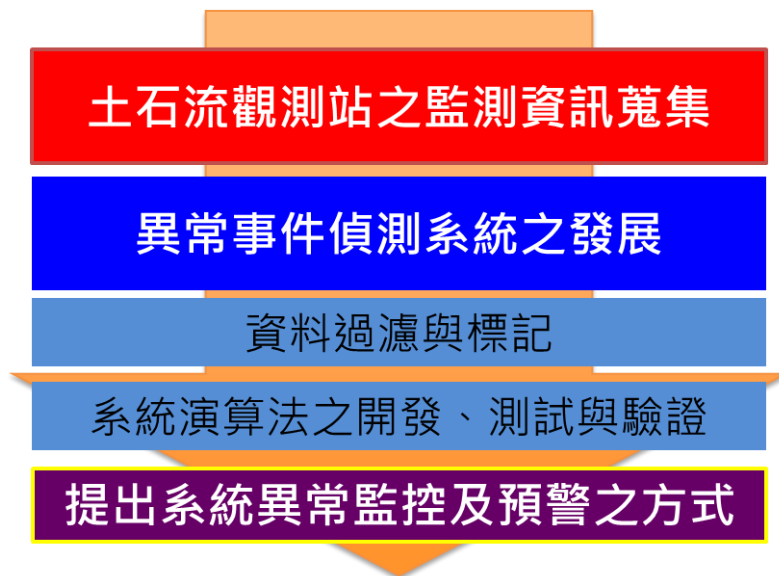


圖 2-1 計畫架構流程

第一節 土石流觀測站之各式監測儀器說明

本研究以陳有蘭溪流域為研究範圍，而在區內有上安、郡坑、豐丘及神木村等四個土石流自動化觀測站，茲將個別測站的監測對象、儀器及位置等說明如表 2-1~表 2-4 所示。而相關監測儀器的用途與資料說明則如表 2-5 所示。

表 2-1 上安土石流觀測站之基本資訊

行政區域	南投縣水里鄉上安村	潛勢溪流編號	投縣 DF165
流域名稱	濁水溪流域	溪流名稱	三廊坑溪
土石流警戒基準值	250 mm	溪流災害類型	溪流型土石流
溪流長度	6.89 km	集水區面積	346 公頃
地質	古第三紀變質岩	上游發生區坡度	30~50°
集水區內崩塌規模	小規模崩塌，崩塌率 > 5%	下游堆積材料情形	土石流材料平均粒徑 > 12"
集水區內植生生長狀況	自然林，植被密集	危害方式	淤埋、撞擊
工程設施	護岸	評定處理順序等級	高
觀測站高程	828 m	觀測站坐標 (TWD97)	X 坐標：237412 Y 坐標：2624605
保全對象	住戶	公共建築	交通設施
	5 戶以上	有(活動中心)	台 21 線、上安橋

摘錄自土石流防災資訊網

表 2-2 郡坑土石流觀測站之基本資訊

行政區域	南投縣水里鄉郡坑村	潛勢溪流編號	投縣 DF167
流域名稱	濁水溪流域	溪流名稱	二廊坑溪
土石流警戒基準值	250 mm	溪流災害類型	溪流型土石流
溪流長度	2.517 km	集水區面積	158 公頃
地質	古第三紀變質岩	上游發生區坡度	> 50°
集水區內崩塌規模	明顯大面積崩塌，崩塌率 < 1%	下游堆積材料情形	土石流材料平均粒徑 > 12"
集水區內植生生長狀況	自然林，植被中等稀疏	危害方式	淤埋、撞擊
工程設施	無	評定處理順序等級	高
觀測站高程	647 m	觀測站坐標 (TWD97)	X 坐標：236766 Y 坐標：2626219
保全對象	住戶	公共建築	交通設施
	5 戶以上	有(活動中心)	台 21 線、郡坑橋

摘錄自土石流防災資訊網

表 2-3 神木村土石流觀測站之基本資訊

行政區域	南投縣信義鄉神木村	潛勢溪流編號	投縣 DF199、投縣 DF226、投縣 DF227
流域名稱	濁水溪流域	溪流名稱	出水溪、愛玉子溪、霍薩溪
土石流警戒基準值	250 mm	溪流災害類型	溪流型土石流
溪流長度	5.518km	集水區面積	866 公頃
地質	新第三紀沉積岩	上游發生區坡度	30~50°
集水區內崩塌規模	明顯大面積崩塌，1% ≤崩塌率≤5%	下游堆積材料情形	土石流材料平均粒徑 介於 3" ~12"
集水區內植生生長狀況	自然林，植被中等稀疏	危害方式	撞擊、漫流改道
工程設施	無	評定處理順序等級	高
觀測站高程	1187 m	觀測站坐標 (TWD97)	X 坐標：235367 Y 坐標：2602749
保全對象	住戶	公共建築	交通設施
	5 戶以上	學校	道路、橋梁

摘錄自土石流防災資訊網

表 2-4 豐丘土石流觀測站之基本資訊

行政區域	南投縣信義鄉豐丘村	潛勢溪流編號	投縣 DF190
流域名稱	濁水溪流域	溪流名稱	陳有蘭溪支流
土石流警戒基準值	250 mm	溪流災害類型	溪流型土石流
溪流長度	2.271 km	集水區面積	162 公頃
地質	古第三紀變質岩	上游發生區坡度	30~50°
集水區內崩塌規模	明顯大面積崩塌，崩 塌率 > 5%	下游堆積材料情形	土石流材料平均粒徑 > 12"
集水區內植生生長狀況	自然林，落石堆積	危害方式	淤埋、撞擊
工程設施	防砂壩、沉砂池、整 流工	評定處理順序等級	高
觀測站高程	633 m	觀測站坐標 (TWD97)	X 坐標：237443 Y 坐標：2618257
保全對象	住戶	公共建築	交通設施
	5 戶以上	有(豐丘國小)	台 21 線、豐丘橋

摘錄自土石流防災資訊網

表 2-5 土石流觀測站之監測儀器說明

儀器	說明	監測頻率
雨量計	量測觀測站當地雨量，作為發佈土石流警戒之重要依據。	每分鐘 1 筆
鋼索檢知器	在易發生土石流的河段架設鋼索檢知器，當土石流沖斷鋼索時，系統自動發生鋼索斷裂通知簡訊。	每分鐘 1 筆
CCD 攝影機	觀測溪流現場動態情況，掌握現場即時影像資訊。	10~30 FPS
地聲檢知器	藉由地聲檢知器量測土石流運動時所引發之地表震動，振動頻率超過警戒值時系統自動發出地聲警戒通知簡訊。	每秒 500 筆
流速計	量測固定河道斷面之水體流速，其原理為水流之表面波紋隨流動時所產生的速度視為水流表面速度。可搭配同一斷面之水位計進行流量之計算。	每秒 1 筆
雷達波水位計	利用雷達波反射計算出水位高低差，長期觀測現地水位之變化。	每秒 1 筆
土壤含水量計	可記錄土壤的水分飽和程度，提供不同土壤及含水量間相關性分析研究之用。	每分鐘 1 筆

第二節 異常事件偵測系統之發展

本研究論之土石流警戒系統的精進，係為提升土石流發生事件的預測準確性。在此異常事件偵測系統的開發過程中，首先將針對訊號係為「正常」或「異常」進行分類，如訊號反應呈「異常」狀態，則進一步依據監測站操作紀錄，釐清該異常狀態的成因並嘗試進行資料標記。然而，此一部分之標記有賴於監測站完整之操作紀錄，始可得到較為確實與完整的結果。最後，透過資料過濾、標記及機器學習的過程，希冀能達到精進土石流警戒系統之最終目的。茲分別將模式相關之演算法、資料處理、過濾與標記等之相關內容說明如下。

一、異常偵測演算法之發展

在異常偵測演算法的方面，本研究提出利用主成分分析（principal component analysis）對異常點較敏感之特性來進行異常偵測，意即異常點的存在會導致資料的主成分方向（principal direction）有很大的偏移。其概念為逐一檢查每一筆資料對主成分方向影響大小來對判斷資料點的異常程度。例如圖 2-2，首先對計算出的資料 PCA 主要方向如圖 2-2 中之綠色線，如果我們移除或者增加一個異常點，再重新計算 PCA 的主要方向時，前後兩個 PCA 主要方向將會產生較大的夾角。換句話說，我們可以透過以測量資料點加入或者移除對 PCA 主要方向變化之影響為基礎來達到異常偵測的目的。除此之外，由於我們假設正常點個數遠多於異常點，如此在移除異常點時，對資料的主要方向角度改變的量可能並不明顯。因此，我們也利用 over-sampling 的方式，多複製一些同樣的點加入資料集，來強化異常點對主成分方向的影響，進而提高異常偵測的準確性。例如圖 2-3，如果我們複製的資料點是正常點，那麼我們重新計算 PCA 的主要方向應該不會有什麼變化；但是如果複製的資料點為異常點，那重新計算的 PCA 主要方向改變量將會更明顯。這個想法的核心概念是我們要把正常點和異常點的影響差距放大，讓我們可以更清楚地辨識那些點是正常點，而那些點是異常點。

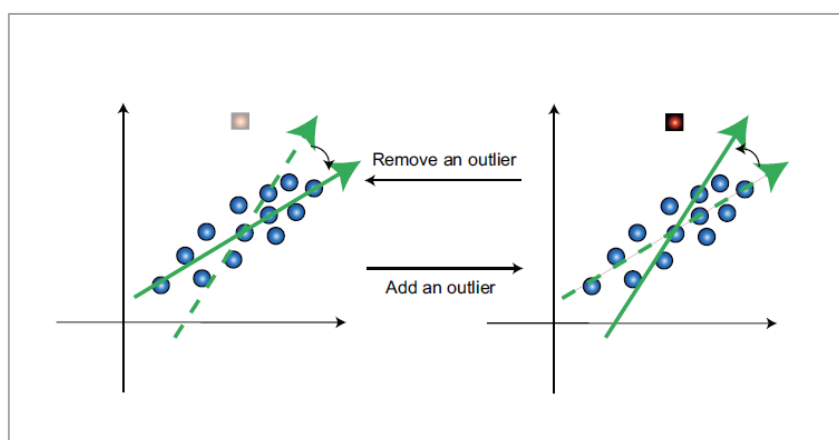


圖 2-2 加入或者移除異常點對主成分方向的影響

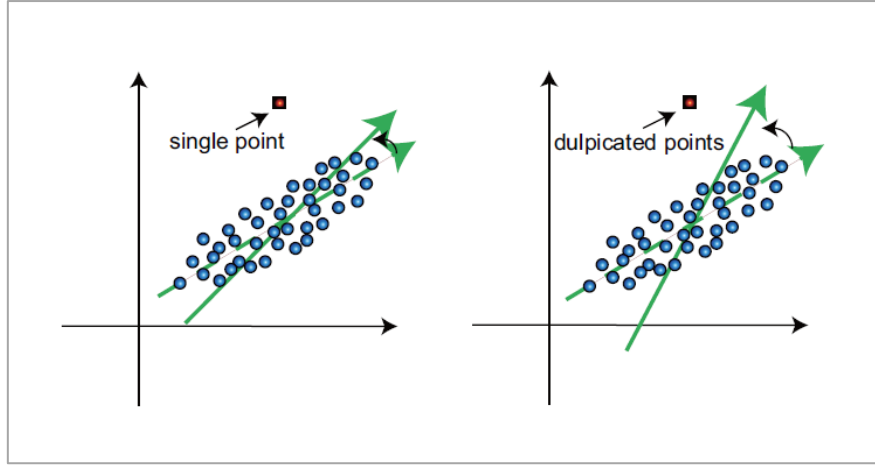


圖 2-3 利用複製異常點加大對主成分方向的影響

透過這些觀察之後，我們得知可以利用 PCA 來做異常偵測，然而在計算 PCA 時需要解下面的最佳化問題：

$$\min_{U \in \mathbb{R}^{p \times k}, \|U\|=I} J(U) = \sum_{i=1}^n \|(x_i - \mu) - UU^T(x_i - \mu)\|^2$$

其中， x_i 為第 i 的資料點， U 為欲求之主成分方向。此問題可推導成求特徵值與特徵向量之問題(eigenvalue decomposition)，其複雜為 $O(p^3)$ ，其中 p 為資料之維度，除了複雜度之外，也令需儲存 $O(p^2)$ 之共變異數矩陣(covariance matrix)。因此為了符合巨量資料運算之需求，我們亦希望透過線上即時更新(online updating)之技巧，達到快速計算 PCA 之目的。為了達到這個問題，我們將先把 PCA 問題轉成最小平方方法問題(least squares problem)如下：

$$\begin{aligned} \min_{U \in \mathbb{R}^{p \times k}, U^T U = I} J_{ls}(U) &= \sum_{i=1}^n \|\bar{x}_i - UU^T \bar{x}_i\|^2 \\ &= \sum_{i=1}^n \|\bar{x}_i - Uy_i\|^2, \end{aligned}$$

其中 U' 是 U 的逼近解， $y_i = U'^T \bar{x}_i$ 。透過把 PCA 化成最小平方方法問題後，我們將採用遞迴最小平方方法問題(recursive least squares)來達到即時更新模型之目的。除此之外，由於我們必須使用 over-sampling 之技術來放大異

常點的影響，因此我們也必須將 over-sampling 概念加入我們求 PCA 主成分方向的遞迴最小平方法問題，其概念為重覆增加欲檢測的資料點，其最佳化問題可以表示如下：

$$\min_{\tilde{\mathbf{U}} \in \mathbb{R}^{p \times k}, \mathbf{U}^T \mathbf{U} = \mathbf{I}} J_{ls}(\tilde{\mathbf{U}}) \approx \sum_{i=1}^n \|\bar{\mathbf{x}}_i - \tilde{\mathbf{U}} \mathbf{y}_i\|^2 + \tilde{n} \|\bar{\mathbf{x}}_t - \tilde{\mathbf{U}} \mathbf{y}_t\|^2$$

其中， n 為複製之次數。透過這樣最佳化問題，在計算上我們將可以應用即時更新的技巧在主成分分析的計算上，除了可以避免高複雜度的矩陣分解並提高計算速度之外，利用及時更新的技巧也無需儲存傳統計算主成分分析所需的共變異數矩陣。另外，透過計算速度上的優勢，我們的方法也可以即時更新主成分方向與偵測異常點。意即新的資料點進來時，若認定是異常點，從資料集中移除該點並發出異常點訊息；反之若為正常點，則加入該點至正常點資料集中，即時更新主成分方向。也因此我們將預期可以應用此具及時更新之非監督式異常偵測模型於異常事件偵測。

二、時間序列資料之特徵學習

由於環境因子即時監測資料大多以時間序列的形式存在，例如感測器可能每一分鐘會感測一次含水量，這些感測到的度量於時間前後有會相關性。傳統上我們可以將每一個時間戳記(time stamp)當作是一個維度，將時間序列以向量的方式呈現，例如圖 2-4 中，時間序列 C 可以用 \mathbf{X} 這個 N 維向量來表示，其中 t_i 為第 i 個時間點。將時間序列藉由每個時間戳記當作一維度的方式轉成向量之後，即可使用歐幾里得距離(Euclidean distance)來計算兩個時間序列之相似度，其方式為計算兩個時間序列各個相對維度差距之總和，愈高代表兩個時間序列愈不相似。然而，利用歐幾里得距離來計算時間序列相似度時，在時間的對齊上要求過於嚴格，也就是若兩個形狀很類似的時間序列，但是因為時間上沒有對準，所計算歐幾里得距離的差就反而比較大。

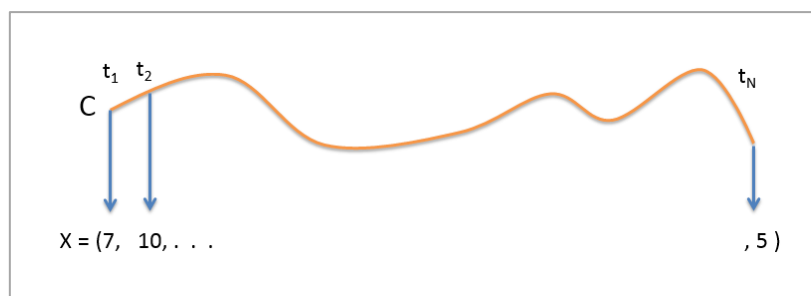


圖 2-4 傳統之時間序列向量化之方式
(將每一個時間戳記當作是一個維度)

有別於傳統的時間序列之表達方式，我們將利用詞袋模型(Bag-of-Words)來對時間序列作特徵學習，以提升負載預測之效能。由於時間序列資料類內變化較大，因此使用傳統的表達資料方式將不適合於許多資料分析的任務上，在這個情形下，我們預期透過多個局部模型(local pattern)來描述時間序列將可以克服這個問題。為了取得這些代表性的局部模型，在給定子序列之長度的情況下，我們首先將藉由滑動視窗的方式從時間序列訓練集取出其所有子序列(如圖 2-5)。

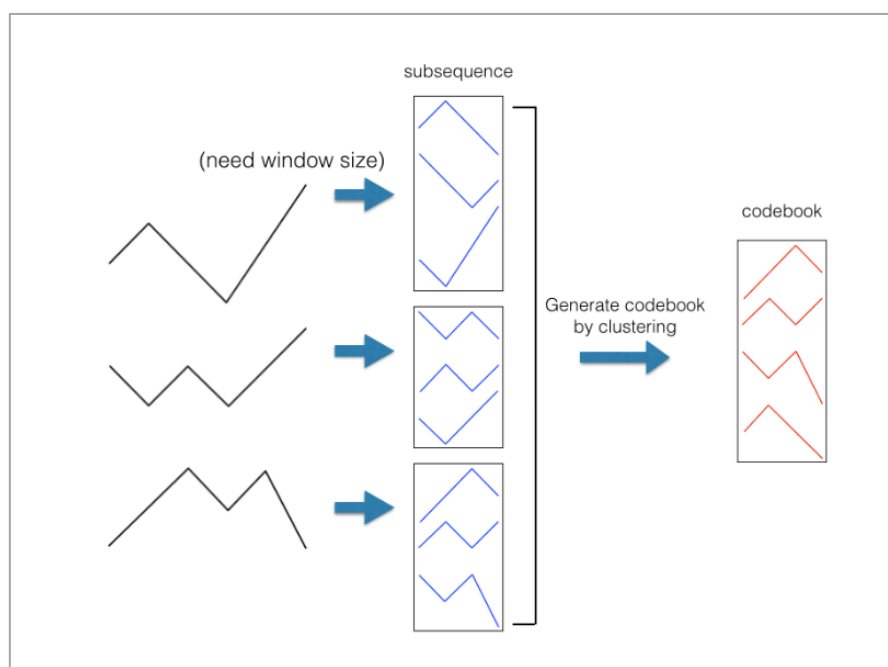


圖 2-5 生成字典之示意圖

因此假設一時間序列長度為 P ，滑動視窗長度為 w (意即子序列之長度)，則每筆時間序列資料將可以產生 $P-w+1$ 個子序列，若有 n 筆資料的話，就可以產生 $n(P-w+1)$ 子序列。一旦獲得這些子序列之後，我們可以利用群聚演算法(例如 k-means)對來這些子序列取出 d 的群代表，這些群代表的集合可視為字典($D \in \mathbb{R}^{w \times d}$)，其每一個群代表可視為字典裡面的字，也是上述的局部模型(如圖 2-5)。

接下來我們將可利用所學習之字典來描述一筆時間序列之資料，也就是利用這些局部模型來描述時間序列之資料。以詞袋模型應用在文件探勘中，我們通常都是文章中出現字的分布代表一篇文章。同樣的，我們也想利用這些局部模型的分布來描述時間序列資料，其示意圖如圖 2-6。

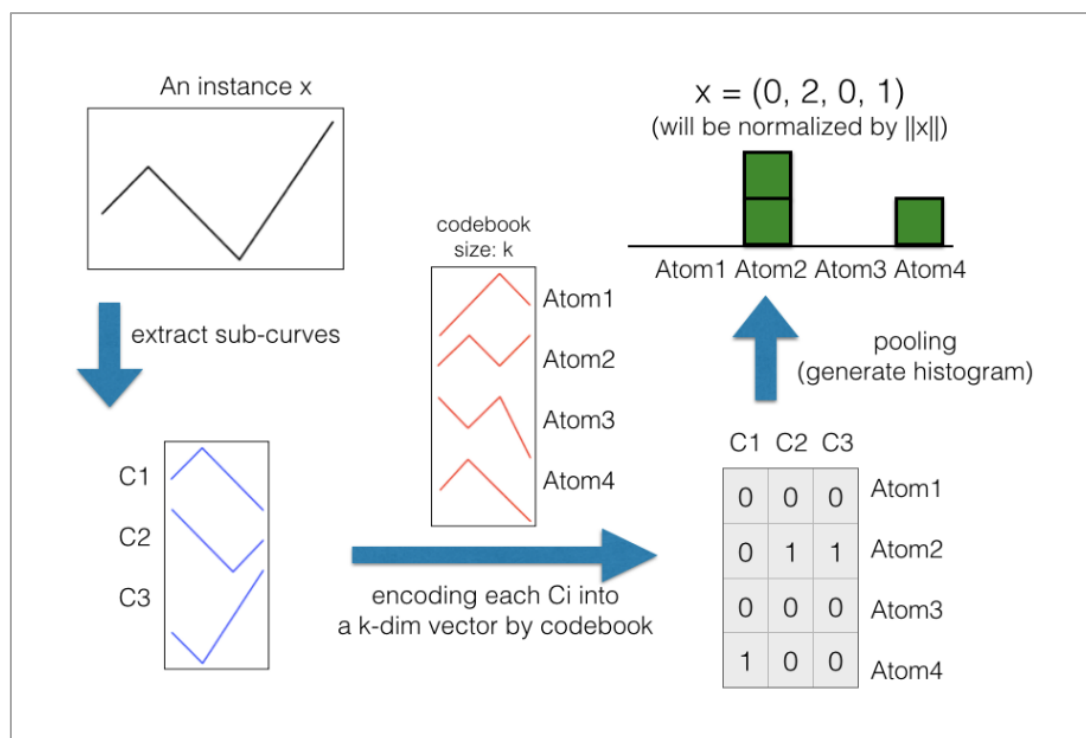


圖 2-6 利用字典來對時間序列作編碼之示意圖

首先，我們對一個時間序列取出 $P-w+1$ 子序列(也就是 $c_1, c_2, \dots, c_{P-w+1}$)，為了整合這些子序列，我們先講每個子序列(c)編碼成 d 為之向量(定義為

$h = (h_1, h_2, \dots, h_d)$ ，其中 d 字典之大小，也就是有幾個局部模型。其中編碼的方式如下式。

$$h_i = \begin{cases} 1 & \text{if } \mathbf{c} \text{ is closest to } i\text{th centroid,} \\ 0 & \text{others.} \end{cases}$$

舉例來說，在字典大小為 5 的情況下，假設一子序列跟第三個群代表最近，則此子序列則會編碼成(0, 0, 1, 0, 0)。一旦我們對 $P-w+1$ 子序列編碼完之後，我們將可聚集這些編碼並歸納出此筆時間序列資料其局部模型的分布，如圖二，最後此筆時間序列資料之向量表示為(0, 2, 0, 1)。其中要注意的是，字典的學習並不局限於群聚演算法，我們也可以利用稀疏編碼(sparse coding)與字典學習法(dictionary learning)做更精緻的時間序列之向量化。

最後利用序列資料特性把序列資料轉成向量後，所以我們將可應用預測模型於時間序列資料之上，其架構圖如下。

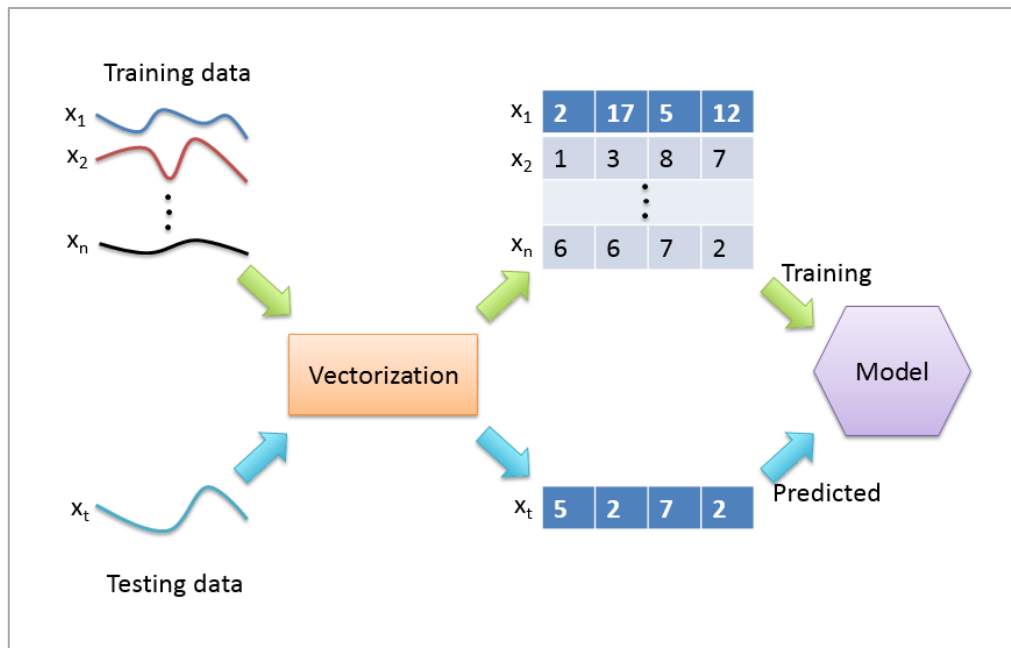


圖 2-7 利用序列資料特性作向量化之資料訓練與預測架構圖

三、資料過濾與標記

異常偵測之目的在於反應系統所監測的數據是否處於非一般的情況，往往異常之訊號代表有意義的活動資訊。而這些有價值的資訊隱藏在大量資料中，單憑人力處理或監控有著容易遺漏，甚至無法處理的問題。另外，過去長期的防災資訊中，由於資料量大也會造成資料標記上的困難，因此如何有效的標記特殊事件，亦為極具挑戰的工作。

有鑑於此，本研究利用異常偵測模型的特性，首先過濾出資料中特殊事件的候選，再做進一步之標記。透過這樣的資料過濾與標記方式，將獲得重要資料的標記，除了可以幫助我們進行監督式(supervised)的學習之外，亦期待透過有效率的資料標記，能夠獲得更精確的預測模型。

然而，以本研究所面臨的問題而言，土石流發生或人為因素的異常事件發生頻率較為罕見，而原始監測資料方面亦無進行相關標記。因此，在使用異常偵測演算法上，非監督式(unsupervised)的異常偵測演算法可能將是實務上比較適合的選擇。此外，能夠處理大量資料的運算亦是在挑選合適的異常偵測演算法條件之一。緣此，本研究採用以 online over-sampling PCA 為基礎之異常偵測模型來進行資料過濾與標記(技術細節如前述)。實際操作流程方面，首先將監測資訊如降雨量、地聲或土壤含水量等原始資料，透過資料特徵學習的技術，表達成適合異常偵測模型之型式。接著再利用轉換過的資料建立非監督式異常偵測模型，並將資料透過此模型，篩選出異常之資料，意即特殊事件之候選。如可利用所計算出的異常值篩選出前 1%可疑之資料後，再以觀測站之操作紀錄為基礎，進一步進行人工標記，將真正之特殊事件挑出。整個過程將透過異常偵測模型過濾掉大量正常資料，這對於第二階段的人工標記中，將大大減低標記人員的工作負擔，進而實現原本無法做到的資料標記。最後透過已標記的資料，亦將可採用更多元的資料分析演算法，進而獲得更好的預測效果。

第三節 異常偵測模型之架構與設定

首先，本研究以每兩小時的時序性降雨資料作為異常偵測模型的輸入值，意即每一個時刻的前兩小時降雨資料為一筆輸入模型的資料。而在資料處理方面，分為兩組，一為時序性資料經過詞袋模型(Bag-of-Words)的重新解讀；二則為原始的時序性降雨資料。藉以比較資料處理方式的精進，是否有助於提升土石流預測之準確性。在模型建構完成後，測試資料首先經過土壤含水量門檻值的過濾，若該時刻前 12 小時的平均土壤含水量大於 80%，始進一步輸入模型進行預測，整體架構如圖 2-8 所示。其中，在圖 2-8 上半部的模型訓練階段，目前係以前一年度的所有降雨資料作為訓練組；在圖 2-8 下半部的測試階段，則以當年度的所有降雨資料進行測試，藉以瞭解異常偵測模型的準確性。如以 2010 年資料進行訓練、2011 年資料進行測試；2011 年資料進行訓練、2012 年資料進行測試。

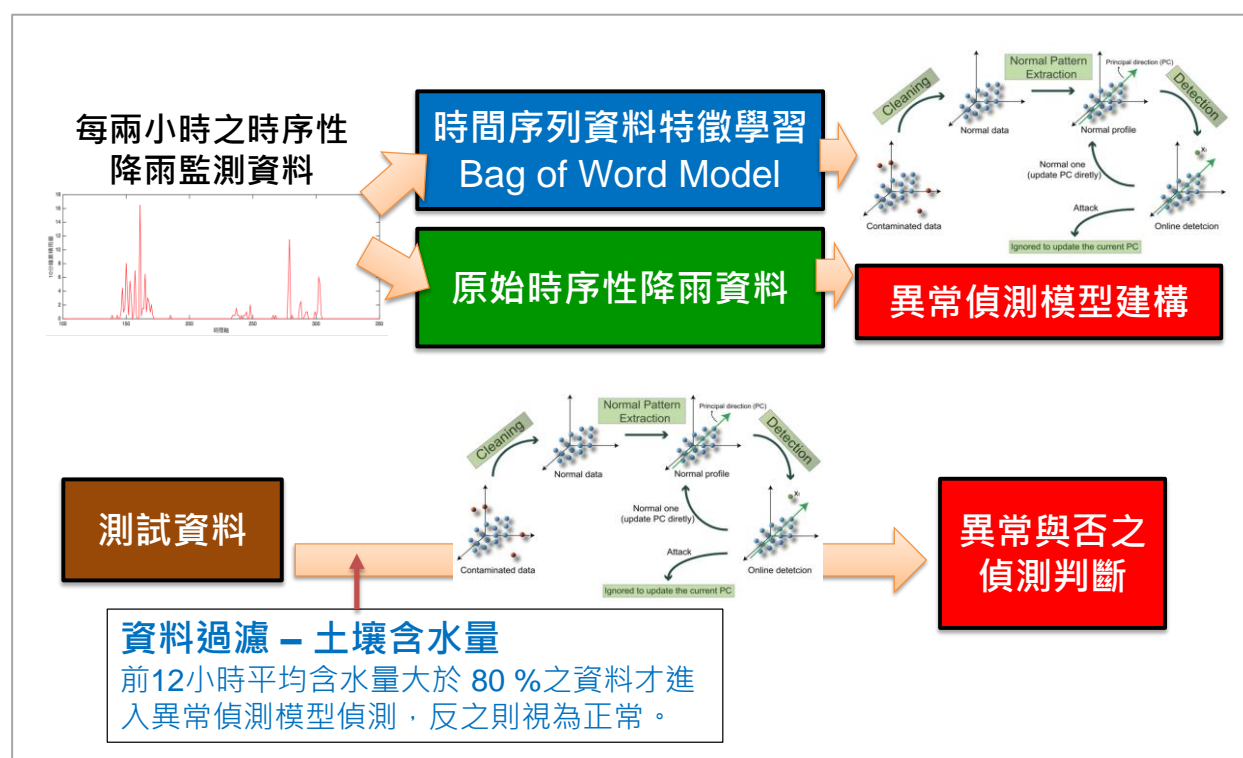


圖 2-8 土石流發生之異常偵測模型架構

本研究所建構之土石流發生異常偵測模型，係為每 10 鐘均根據該時刻前兩個小時的降雨資料進行是否異常的偵測。當模型偵測到異常時，系統即發出紅色警戒；在發出紅色警戒後，若持續 12 小時系統沒有偵測到異常，即轉為黃色警戒；而在黃色警戒後，再持續 12 小時系統皆無偵測到異常，即轉為綠燈(解除警戒)。整體發布與解除警戒之架構如圖 2-9 所示。

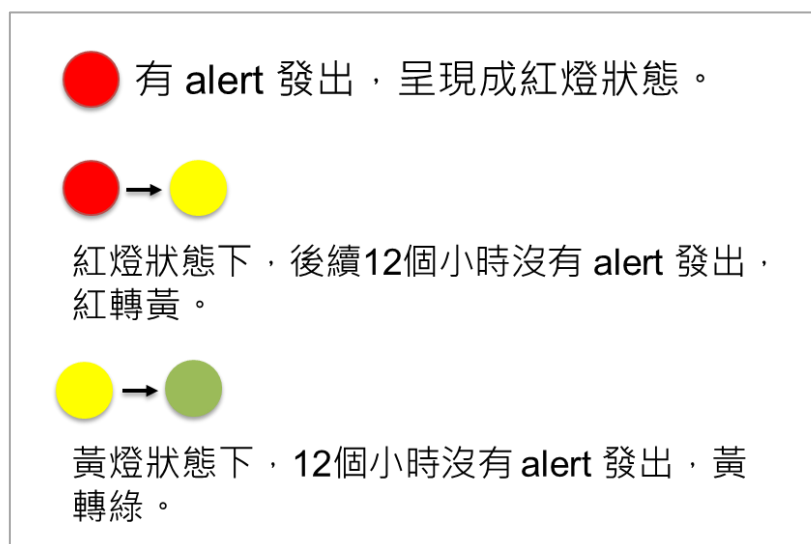


圖 2-9 異常偵測模型之警戒發布與解除架構

第四節 模式評估

在發展異常偵測模式後，本研究採用命中率與可信度等指標來評估模式的可行性，並和現行之警戒模式作綜合性的比較，以觀察本模式之效用。

一、命中率與可信度

為量化由模式產生之警報的可行性評估，本研究採用命中率及可信度進行計算。其定義分別說明如下：

命中率=土石流發生前有發出預警次數/土石流發生事件數

可信度=土石流發生前有發出預警次數/總預警次數

由命中率可以得知模式在土石流發生前發布警戒的準確性，命中率愈高表示模式愈能在事前偵測到土石流事件並發布警戒，而可信度則可得知模式產生預警的誤報程度，可信度愈高表示誤報率愈少。

二、水保局有效累積雨量及紅色警戒

為探討此模式應用於不同測站之可行性，因此，本研究將對照真實土石流事件，並與現行水保局有效累積雨量計算方式所發出之警戒、歷年水保局所發佈之紅色警戒資訊，進行綜合分析及歸納。下圖 2-10 為現行土石流警戒發佈之運作方式。

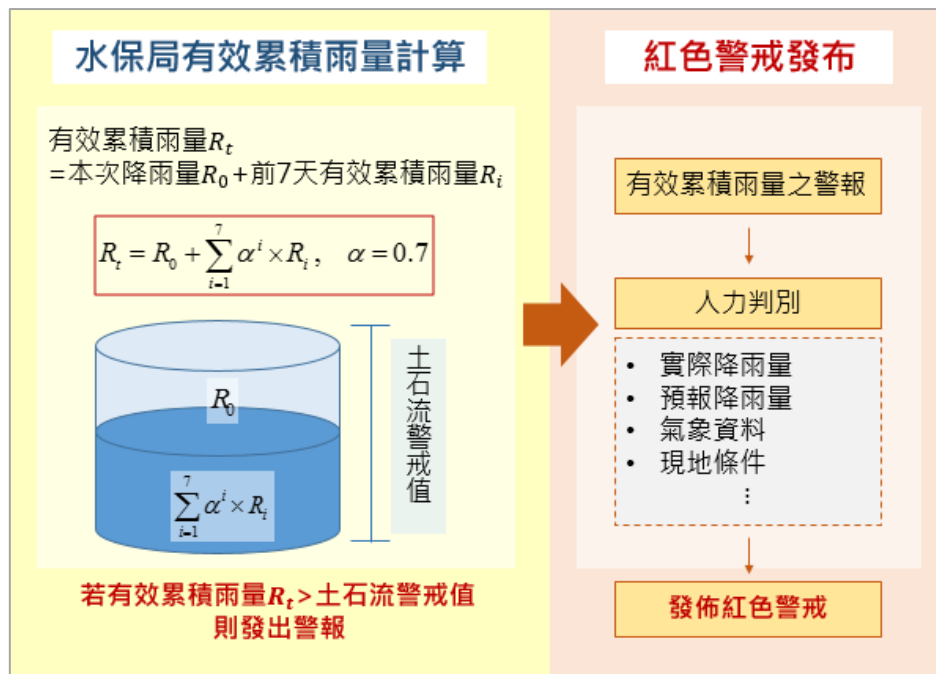


圖 2-10 水保局有效累積雨量方法及紅色警戒發布模式

首先，水保局計算有效累積雨量 R_t 的方式為當場降雨 R_0 加上前七天之有效累積雨量，並考量降雨日衰減係數 α 令有效累積雨量依時間進行指數的折減，如前一日的有效累積雨量乘上 α^1 、前二日則乘上 α^2 ，以此類推至前七日，再進行累加。其中，水保局所採用之衰減係數 α 為 0.7。該計算公式可表示為：

$$R_t = R_0 + \sum_{i=1}^7 \alpha^i \times R_i, \quad \alpha = 0.7。$$

當有效累積雨量大於土石流警戒基準值時，便發佈土石流警報。故本研究亦採用相同的方式及衰減係數，計算各測站之有效累積雨量，並且依水保局所訂定之土石流警戒基準值發布警報。

而在紅色警戒發布方面，則是以有效累積雨量所發出警報為基礎，若實際降雨已達土石流警戒基準值，需再輔以氣象資料加以判讀後，以人為決策的方式發布紅色警戒。

第三章 結果與討論

依前所述，本研究蒐集土石流觀測站之監測資訊後，將進行各監測因子的探討，以釐清各因子與土石流發生機制之間的關係，並且挑選出適用於土石流預警模式的因子，進行後續模式之操作。接著再將該因子以詞袋模型進行試驗，探討該模式的可行性。最後再以該模型為基礎，建立土石流警戒系統。

第一節 不同監測因子之測試探討

一、土壤含水量

首先，利用神木村 2010 年至 2012 年土壤含水量的時序性監測資料進行觀察。結果發現，在連續降雨期間，神木村站所埋設的土壤含水量計，其值皆接近飽和狀態，特別是在雨量較小或是雨量較大時，土壤含水量的變異不大，由圖 3-1 之範例可知，降雨量較小時，土壤含水量約介於 86 % 至 88 % 之間；而在降雨量較大時，土壤含水量亦介於同一範圍之內。

由土石流發生的機制可知，土石流發生時一定伴隨著降雨。而透過土壤含水量的觀察，發現其在連續降雨期間的變異不甚顯著。因此，以神木村站所埋設的土壤含水量計而言，或因其埋設位置的影響，導致受降雨的影響較不顯著。後續仍待其他土石流觀測站的測試結果，始可得到較完整的探討。

然而，如從另一個面向探討土壤含水量於土石流發生偵測之應用，假設土石流發生時間皆處於連續降雨期間。因此，在土石流發生時的土壤含水量一定是接近飽和之狀態。透過神木村站實際觀察數據後亦發現，實際土石流事件發生時的前 12 小時平均土壤含水量，皆高過 80 %。

透過此一觀察現象的解讀，或可嘗試將土壤含水量應用於資料的過濾，意即每一時刻的前 12 小時平均土壤含水量若超過 80 %，始進行土石流發生與否的預測，若該時刻的前 12 小時平均土壤含水量未超過 80 %，則視為正常資料，且不進行是否發生土石流的預測。

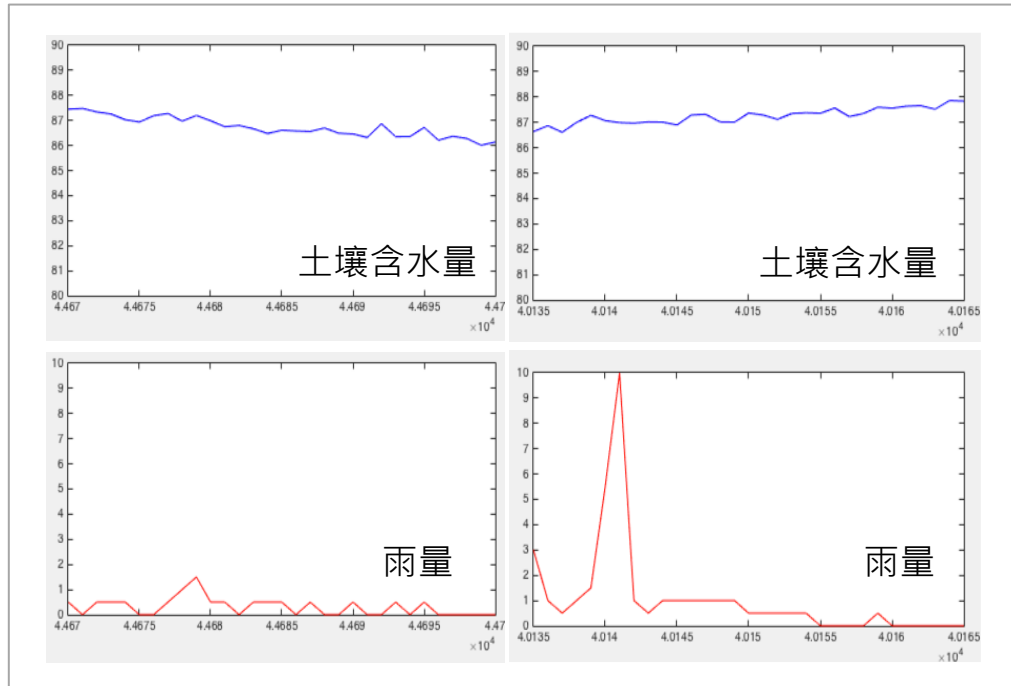


圖 3-1 土壤含水量及雨量關係之示意圖

二、地聲

透過地聲相關文獻的回顧以及神木村實測資料的解讀，可發現地聲訊號確實能有效的反應土石流的發生。顯然土石流發生時，河床的振動頻率與平時正常狀況的振動頻率有所差異。舉例而言，神木村站於 2012 年 5 月 4 日 15 點 56 分發生一場土石流事件，其地聲訊號即有明顯的突起，如圖 3-2 所示。

然而，透過資料的觀察亦可得知，地聲訊號的反應時間，與土石流發生時間極為相近。因此，若欲使用地聲訊號進行土石流發生與否的預測，在災前預警時間方面，可能無法獲得較長的應變時間。此點是為目前利用地聲訊號進行土石流發生預測方面的問題所在。

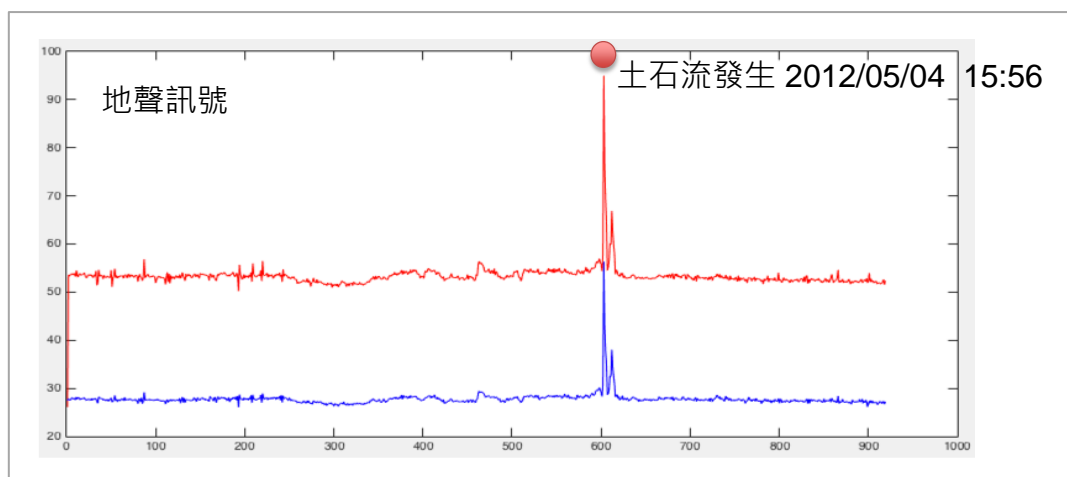


圖 3-2 地聲訊號與土石流發生時間關係之示意圖

三、流速

以土石流發生機制而言，在土石流實際發生前，可能因上游兩側的土石崩落而堆積、阻斷河道，使下游河道的流量(速)有下降之現象。由此假設可知，流速監測值的變化應可一定程度反應土石流的發生與否。以神木村站 2011 年 07 月 19 日 03 點 19 分的土石流發生事件而言，在土石流發生的前 10 分鐘左右，流速觀測值有明顯下降的現象，如圖 3-3 所示。

然而，目前所取得的流速資料仍不甚完整，未來應將持續探討應用流速進行土石流預測之可行性。

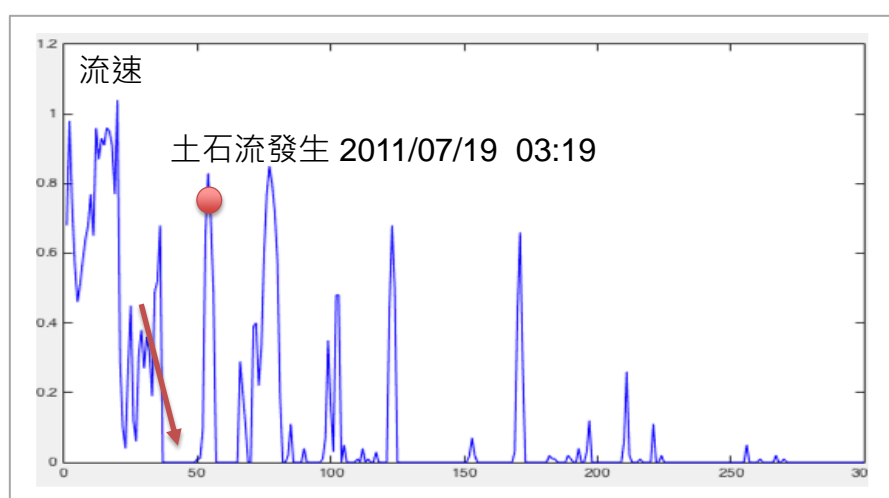


圖 3-3 流速與土石流發生之關係示意圖

第二節 異常偵測模型之初步測試

依照前述異常偵測模型之建構說明，本研究初步以神木村土石流觀測站 2010 年至 2012 年的監測資料進行測試。其中，係以 2010 年資料作為訓練組、2011 年資料進行測試；2011 年資料作為訓練組、2012 年資料進行測試。分別均以兩種資料處理方式進行，最後並將模型判斷結果與真實土石流發生事件進行比對，茲將結果整理如表 3-1 所示。

由表 3-1 的測試結果可知，首先以使用原始時序性降雨監測資料的組別而言，模型在 2011 年至 2012 年間，總計發布 25 次的土石流發生警戒，其中 5 次真實土石流發生事件均有準確的判斷；同時亦有 20 次的誤報情形。再者，以特徵學習法的組別而言，模型在 2011 年至 2012 年間，總計發布 14 次的土石流發生警戒，其中 4 次真實土石流發生事件均有準確的判斷，而有一場土石流發生事件，模型並無發出警戒；同時亦有 10 次的誤報情形。

表 3-1 神木村觀測站之模型測試結果一覽

組別		Alert	False Alert	True Alert
2011	原始時序性降雨監測資料	10	7	3
	時間序列資料特徵學習法 Bag of Word Model	5	3	2
2012	原始時序性降雨監測資料	15	13	2
	時間序列資料特徵學習法 Bag of Word Model	9	7	2
總計	原始時序性降雨監測資料	25	20	5
	時間序列資料特徵學習 Bag of Word Model	14	10	4

緣於上述初步測試結果可知，透過時間序列資料特徵學習處理的組別，能有效降低模型發布警戒的誤報率。然而，確有一場真實的土石流發生事件，模型並無發出警報。本研究進一步探討此兩種不同的資料前處理

方式對模型所產生的影響，首先在原始降雨資料的模型判斷方面，其所偵測到的異常行為，該時段內雨量的變化較傾向瞬間的降雨；而在特徵學習處理的組別內，其所偵測到的異常行為則較偏向較長時間的連續降雨，如圖 3-4 所示。若以實際土石流發生的機制而言，需要有充沛的流量，始較有可能引發土石流的發生。因此，就兩種不同資料處理方式所偵測到的異常行為而言，應是以特徵學習法的組別，所偵測的較長時間連續降雨，較符合實際土石流發生的狀況。

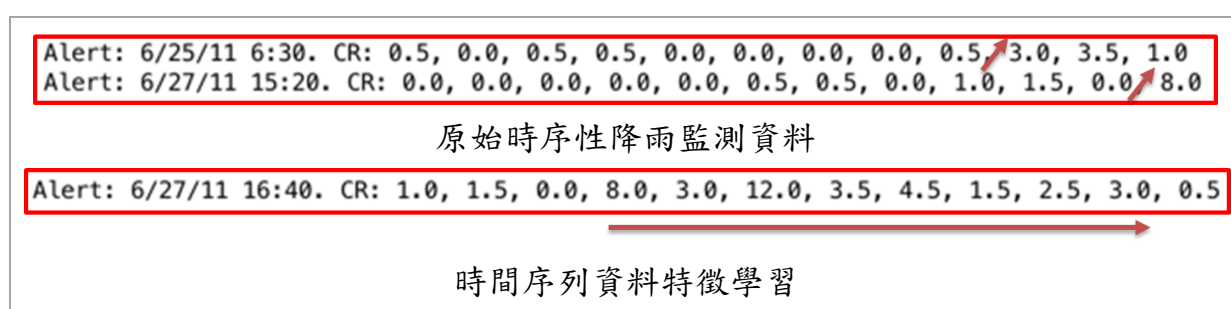


圖 3-4 不同組別偵測之異常行為特徵

此外，在特徵學習法的模型判斷中，有一場真實土石流事件並無發出警報。透過該場事件的檢視發現，土石流發生時間點的前 12 小時，其累積雨量僅約 10.5 mm，如圖 3-5 所示。此應為即為特殊的土石流事件，若將此事件的雨量特徵視為一般土石流發生時的雨量特徵，可能將造成誤報率大幅提升的現象。此一結果亦說明，使用原始資料進行模型判斷時，誤報率有較高的現象。

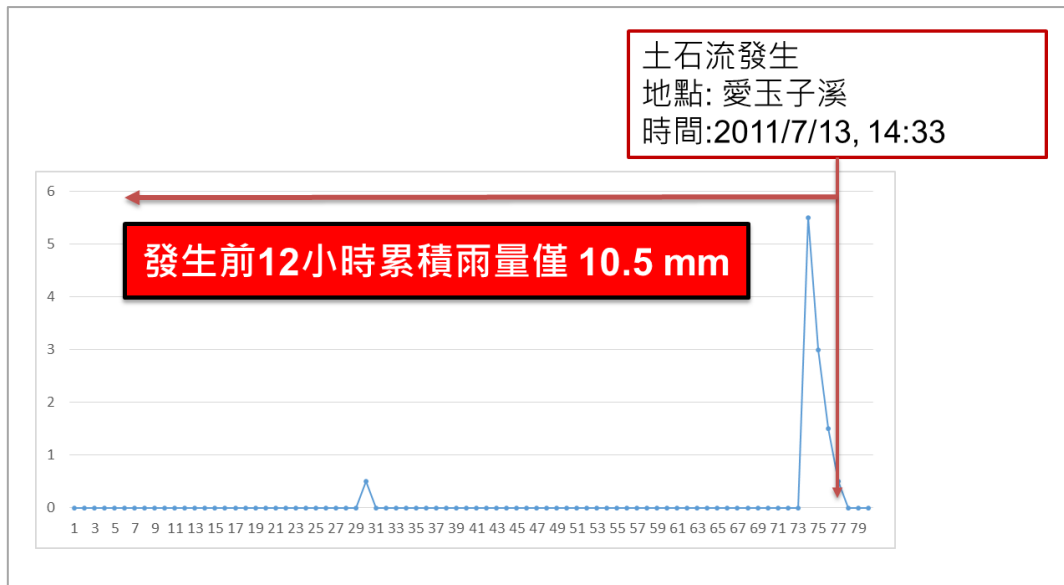


圖 3-5 特徵學習法模型之漏報事件

由異常偵測演算法與詞袋模型初步應用於神木村觀測站之結果得知，該模型所偵測到的異常行為較符合實際土石流發生的情況，其警報的準確性不僅高於原始降雨資料的模型，更能在大部分的土石流事件前發出警報。因此本研究將以詞袋模型為基礎建立自動化的控制流程，發展土石流警戒模式，並且分別於神木、豐丘、上安及郡坑等四個測站進行測試，並比較現有之土石流警戒機制。

第三節 土石流預警系統模式建置

本研究以前述所發展之異常偵測演算法與詞袋模型，建置土石流警戒模式。同時，分別應用至神木、豐丘、上安及郡坑等四個測站，並以 2011 年至 2015 年的實測降雨資料進行測試，藉以觀察此警戒模式應用於土石流防災預警的可行性。整體模式操作流程如圖 3-6 所示。

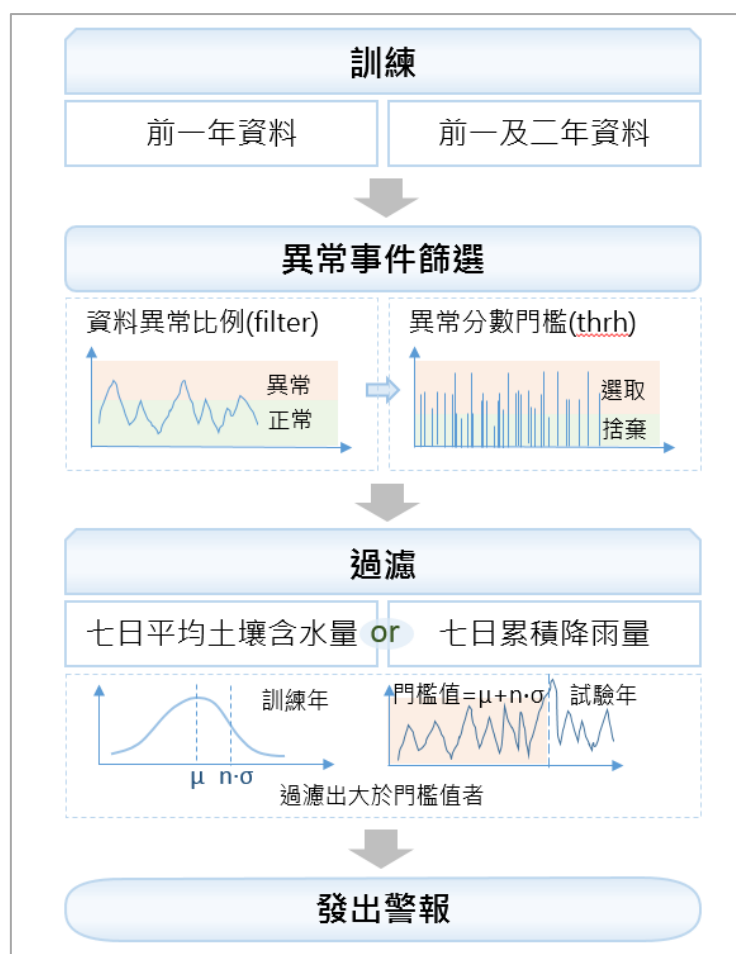


圖 3-6 模式操作流程

由於詞袋模型的訓練資料會影響試驗資料的警戒發佈結果，在初步的測試中，是以測試年的前一年資料作為訓練資料。然而訓練資料的長度是否會直接影響土石流警戒的準確性仍需要進行相關的試驗，因此，本研究除了延續初步試驗之以一年長度為訓練年外(意即 2010 年資料訓練、以 2011 年資料測試)，另外再以測試資料的前兩年資料進行訓練作為對照組(意即 2010~2011 年資料訓練、以 2012 年資料測試)，以此方式來觀察是否訓練年份的長度，將影響警戒發布的準確性。

在以詞袋模型進行訓練前，需先設定特定比例(filter)，將此比例的資料視為會發生土石流的異常事件並進行訓練，在與測試資料比較後，進而得到測試資料的異常值。接著，再以一異常分數門檻比例(thrh)篩選出異常分

數高於此門檻比例的異常事件再發佈警戒。此兩參數的調整為詞袋模型預測準確與否具有關鍵性的影響，可以直接過濾掉模式認定之非異常事件。在此一部分的設定上，考量到災害發生的影響，期先以較為保守的設定，將使模式可以發布所有可能發生土石流的事件。其後，再以較嚴苛且符合物理現象的門檻，來篩除不合理的誤報事件，始發出最終的土石流警報。

而在濾除誤報事件的門檻值參數方面，由監測因子與詞袋模型初步的測試，以神木村觀測站為例，其土壤含水量在土石流事件發生前 12 小時皆大於 80 %。此外，充份的雨量為觸發土石流發生的機制之一，因此亦將累積雨量納入過濾誤報事件的條件。緣此，本研究納入累積雨量和土壤含水量作為過濾誤報事件的門檻，並參考水保局目前以七日為時間長度計算有效累積雨量的方式，設定過濾誤報事件的門檻值參數為七天土壤平均含水量及七天累積降雨量。

因此，本研究根據訓練資料之分佈特性，假設其七日平均土壤含水量分佈為常態分佈；七日累積雨量為對數常態分佈，並分別以平均值加上 n 倍標準差決定門檻值。接著，由詞袋模型所產生的警戒亦分別計算七天土壤平均含水量及七天累積降雨量，若其中一項大於門檻值者，即發布土石流警戒。

第四節 土石流預警系統模式測試結果與討論

因各測站具有不同之地文和水文特性，以及其他儀器觀測之因素，使各觀測站需要有各自不同之參數進行測試。因此，本研究將針對各測站進行試驗，並且找出盡可能準確發布土石流預警之最佳參數，最後再依各測站之預警與真實土石流事件及水保局有效累積雨量和紅色警戒做綜合探討。

一、可信度與命中率分析

將詞袋模型應用至神木、豐丘、上安、郡坑等四個測站，並嘗試多組參數後，得出最合適各測站的參數組合，使詞袋模型可以依各測站的特性發出土石流警報，根據目前測試結果而言，最佳參數設定如表 3-2 所示。

表 3-2 各測站經試驗後之最佳參數設定

測站名	神木	豐丘	上安	郡坑
訓練年長度	1 年	1 年	2 年	2 年
異常資料比例(filter)	0.1	0.05	0.5	0.5
異常分數門檻比例(thrh)	0.025	0.05	0.5	0.5
七日累積降雨量	3 倍標準差	1 倍標準差	2 倍標準差	2 倍標準差
七日平均土壤含水量	3 倍標準差	1.5 倍標準差	3 倍標準差	3 倍標準差

再者，本研究除了提升模式之準確性外，亦將現行的土石流預警方式納入比較，探討詞袋模型應用於土石流預警系統之可行性。本研究對照之預警方式為：由水保局計算有效累積雨量的方式發出之警報，以及水保局實際發布之土石流紅色警戒紀錄。

表 3-3 則為最佳參數組合所發出之警報、由水保局計算有效累積雨量的方式發出之警報，以及水保局發布之土石流紅色警戒之紀錄等三種警戒方式，進行命中率及可信度的綜合比較。

表 3-3 命中率與可信度分析

測站	R	詞袋模型				水保局有效累積雨量				紅色警戒			
		命中率	可信度	C	P	命中率	可信度	C	P	命中率	可信度	C	P
上安	0	-	0.00	0	3	-	0.00	0	5	-	0.00	0	3
郡坑	0	-	0.00	0	1	-	0.00	0	3	-	0.00	0	3
神木	8	0.75	0.46	6	13	0.25	0.25	2	8	0.25	0.29	2	7
豐丘	1	1.00	0.25	1	4	0.00	0.00	0	4	0.00	0.00	0	4

R=土石流發生事件數，C=警戒命中次數，P=總預警次數

命中率=土石流發生前有發出預警次數/土石流發生事件數

可信度=土石流發生前有發出預警次數/總預警次數

由表 3-3 可知，神木站於 2011 年至 2015 年間共發生 8 次土石流事件，在詞袋模型所發出的 13 個警戒中，有 6 次可以命中土石流事件。然而，無論是以水保局計算有效累積雨量的方式或是紅色警戒，僅能捕捉到兩場土石流事件，且可信度與命中率皆低於由詞袋模型發出的預警。

而豐丘站於此期間只發生一次的土石流事件，由詞袋模型發出 4 個預報，可以命中該土石流事件。相較於水保局有效累積雨量和紅色警戒無法於該事件發出預警，詞袋模型能在豐丘站發揮相當大的效用。

二、各測站分析

在綜合比較預警方式後，我們可以得知詞袋模型在決定出各測站之參數後，所發布之土石流預警無論是命中率或是可信度，皆高於水保局有效累積雨量和紅色警戒。而針對各測站之預警特性，將討論如下。

(一) 神木站

在神木站的測試中，初步採用較寬鬆之參數設定進行試驗，以進行前一年資料當作訓練資料以及採用前兩年資料當作訓練資料的比較。此時參數之設定為：異常分數門檻比例(thrh)為 0.025、七日累積雨量及平均土壤含水量以 2 倍標準差為門檻，以及訓練資料中異常的比例定義(filter_r)設定為 0.5，也就是訓練資料的一半判定為異常的判定標準。

此項比較的時間為 2012~2015，這段時間中總共發生 5 場土石流事件。預測結果可從表 3-4 得知，以前兩年作為訓練資料的預測結果，命中率大幅降低，5 個土石流事件中無法遇到其中兩個土石流事件，故選擇採取以前一年資料為訓練資料的方式。

表 3-4 神木站訓練時間長度之結果比較

訓練資料長度	準確預報次數	總預報次數	命中率	可信度
1 年	5	95	1	0.053
2 年	3	42	0.6	0.071

因此，神木站以訓練資料為一年長度進行試驗，並將各參數嘗試各組合進行多次試驗，找出可令土石流預警之命中率及可信度最佳的參數組合。而最終我們所嘗試的組合及各方式之預警時間如下表 3-5 所示。

表 3-5 神木站試驗最佳參數設定與土石流事件及各警報比較

神木站試驗最佳參數設定			
異常資料比例(filter) = 0.1		七日之累積降雨量 = 3 倍標準差	
異常分數門檻比例(thrh) = 0.025		平均土壤含水量門檻值 = 3 倍標準差	
		訓練資料時間長度 = 1 年	
神木站土石流事件與各警報比較			
土石流事件	詞袋模型	有效累積雨量	紅色警戒
2011/07/13 14:30	2011/07/19 03:30	2012/06/10 20:30	2012/06/10 17:00
2011/07/19 23:00	2012/05/03 17:40	2013/05/21 05:40	2012/08/02 14:00
2011/11/10 03:20	2012/05/20 09:30	2013/07/13 06:50	2013/07/13 03:00
2012/05/04 16:00	2012/06/10 12:10	2013/08/22 05:50	2013/08/21 20:00
2012/06/10 19:00	2012/08/02 06:20	2014/07/23 16:30	2014/07/23 14:00
2013/05/19 07:00	2012/08/16 21:50	2015/05/25 02:10	2015/08/08 17:00
2013/07/13 07:00	2013/05/17 05:50	2015/08/09 02:10	2015/09/29 02:00
2014/05/20 12:50	2013/05/19 08:40	2015/09/29 03:20	
	2013/05/21 06:30		
	2013/07/13 02:00		
	2013/08/21 21:10		
	2014/05/20 14:10		
	2015/05/26 15:10		
命中率	0.75	0.25	0.25
可信度	0.46	0.25	0.29

由上表可知，在 2011 年至 2015 年間共發生 8 次土石流事件中，以水保局計算有效累積雨量的方式所發布之 8 次預警僅命中 2 次，而紅色警戒雖發布之次數較少，但也同樣補捉到 2 次的土石流事件。而詞袋模型雖發布了 13 次預警，卻有 6 次可以命中土石流事件，無論是可信度和命中率，皆較其餘兩種方式高。

若我們檢視神木站之土石流事件，可發現詞袋模型所漏報的事件不僅七日平均土壤含水量較低，七日累積降雨量更是不到 100mm，此數值較低的土石流事件，因低於門檻值，無法由詞袋模型發布警報，其事件之降雨量時間序列如圖 3-7 所示。

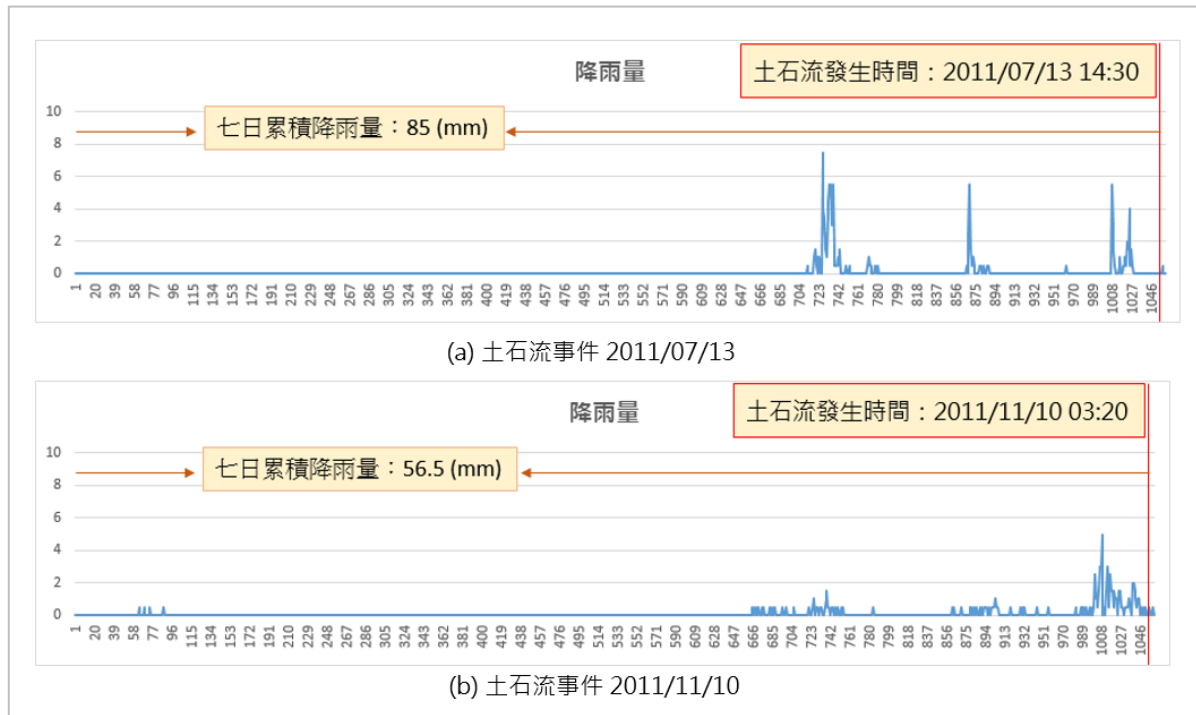


圖 3-7 詞袋模型漏報之土石流事件

(二) 豐丘

豐丘站初步嘗試以前一年資料作為訓練資料與以前兩年做訓練資料作比較，此時參數之設定為：異常分數門檻比例(thrh)為 0.02、七日累積雨量及平均土壤含水量以 1 倍標準差為門檻，以及訓練資料中異常的比例定義(filter_r)設定為 0.05。由表 3-6 之試驗結果可知，用前兩年的資料作為訓練資料，則會無法預警 2012 年的土石流事件，故選擇以前一年為訓練資料的設定進行相關測試。

表 3-6 豐丘站訓練時間長度之結果比較

訓練資料長度	準確預報次數	總預報次數	命中率	可信度
1 年	1	20	1	0.05
2 年	0	14	0	0

因此，豐丘站亦以訓練資料為一年長度進行試驗，並將各參數嘗試各組合進行多次試驗，找出可令土石流預警之命中率及可信度最佳的參數組合。而最終我們所嘗試的組合及各方式之預警時間如下表 3-7 所示。

表 3-7 豐丘站試驗最佳參數設定與土石流事件及各警報比較

豐丘站試驗最佳參數設定			
filter = 0.05		七日之累積降雨量 = 1 倍標準差	
thrh = 0.02		平均土壤含水量門檻值 = 1.5 倍標準差	
		訓練資料時間長度 = 1 年	
豐丘站土石流事件與各警報比較			
土石流事件	詞袋模型	累積有效降雨量	紅色警戒
2012/05/03 16:00	2012/05/03 15:40	2012/06/12 13:40	2012/06/11 20:00
	2012/05/20 09:50	2012/08/02 08:50	2012/08/02 08:00
	2012/08/02 06:50	2013/07/13 07:00	2013/07/12 23:00
	2013/05/19 09:20	2013/08/22 07:20	2013/08/21 20:00
命中率	1.00	0.00	0.00
可信度	0.25	0.00	0.00

豐丘站於 2011 年至 2015 年間僅發生一次土石流事件，在詞袋模型所發出的 4 個警戒中，有 1 次可以命中土石流事件。然而無論是以水保局計算有效累積雨量的方式或是紅色警戒，卻無法警戒出該土石流事件，兩種方式皆產生四筆誤報。

(三) 上安

上安站之試驗中，將七日累積降雨量的門檻值設定為 3 倍標準偏差與 2 倍標準偏差的結果為相同，而七日平均土壤含水量則設定為 3 倍標準偏差。而因為上安站於 2011 年至 2015 年並無土石流事件發生，由表 3-8 訓練資料長度的比較中可知，以訓練資料長度為 1 年所預警的次數多於以 2

年為訓練資料長度所發布的警戒，因此本研究於上安站用誤報次數較少的訓練長度為 2 年之訓練資料繼續嘗試其他相關之參數。

表 3-8 上安站訓練時間長度之結果比較

訓練資料長度	準確預報次數	總預報次數	命中率	可信度
1 年	0	8	0	0
2 年	0	3	0	0

由試驗結果(如下表 3-9)，可得知上安站的訓練長度用兩年所產生的預報次數較少，並且在三次預報中有兩次的預報和水保局所發布的紅色警戒相符，並提早發布，使地方可以及早準備。然而上安站在 2010 至 2015 年間並無土石流的災害發生，若以此準則來檢視各預報，則此三種警報方式的結果皆被視為誤報，而誤報的次數以詞袋模型和紅色警戒最少。

表 3-9 上安站試驗最佳參數設定土石流事件與各警報比較

上安站試驗最佳參數設定				
filter = 0.5			七日之累積降雨量 = 3 倍標準差	
thrh = 0.5			平均土壤含水量門檻值 = 3 倍標準差	
			訓練資料時間長度 = 2 年	
上安站土石流事件與各警報比較				
土石流事件	詞袋模型		累積有效降雨量	紅色警戒
無	2013/07/13	05:20	2010/07/21 16:30	2012/08/02 11:00
	2013/08/21	16:30	2012/08/02 09:50	2013/08/21 20:00
	2013/08/29	10:10	2013/07/01 11:30	2013/08/29 13:00
			2013/08/22 06:10	
			2014/06/07 14:30	
命中率	0.00		0.00	0.00
可信度	0.00		0.00	0.00

(四) 郡坑

此測站試驗時將七日累積降雨量與平均土壤含水量的門檻值各以 2 和 3 倍標準偏差的組合進行測試，因為郡坑站於 2011 年至 2015 年並無土石流事件發生，以訓練資料長度為 1 年所產生預警的次數多於以 2 年為訓練資料長度所發布的警戒(表 3-10)，因此於郡坑站之訓練資料時間長度採用兩年以進行操作。

表 3-10 郡坑站訓練時間長度之結果比較

訓練資料長度	準確預報次數	總預報次數	命中率	可信度
1 年	0	2	0	0
2 年	0	1	0	0

由訓練長度為兩年進行試驗，並嘗試出最佳參數組合後，其所發布之警戒與各預警模式之比較如下表 3-11 所示。由表 3-11 可知，詞袋模型所產生之警戒次數在三者方式之中最少。然而郡坑站在 2010 至 2015 年間亦無土石流的災害發生，若以此準則來檢視各預報，則此三種警報方式的結果皆被視為誤報，其中又以詞袋模型的誤報次數最少。

表 3-11 郡坑站試驗最佳參數設定土石流事件與各警報比較

郡坑站試驗最佳參數設定			
filter = 0.5		七日之累積降雨量 = 2 倍標準差	
thrh = 0.5		平均土壤含水量門檻值 = 3 倍標準差	
		訓練資料時間長度 = 2 年	
郡坑站土石流事件與各警報比較			
土石流事件	詞袋模型	累積有效降雨量	紅色警戒
無	2012/08/02 11:10	2012/08/02 09:50 2013/07/13 11:10 2013/08/22 07:30	2012/08/02 11:00 2013/07/13 11:00 2013/08/21 20:00
命中率	1.00	0.00	0.00
可信度	0.25	0.00	0.00

三、操作結果小結

整體而言，在上安及郡坑站的部分，因歷史上沒有土石流災害，以此準則來檢視，則詞袋模型的誤報次數最低。而在神木及豐丘這兩測站的表現，由詞袋模型所發出的警報，無論是命中率或是可信度，皆大於另外兩種方式。此外，以災害應變階段的實際操作層面而言，紅色警戒的發布係依據有效累積雨量是否超過門檻後，再以人力綜合判斷氣象情資而決定。而本研究所發展的土石流警戒模式，為一自動化之判斷模式，可降低人為判斷的負擔。同時，在命中率與可信度方面，皆較現行警戒發布之情形為準確。因此，未來土石流的警報或許可由現有的詞袋模型再進行發展，以減少人力與判讀時間，並提高土石流警報的準確性。

第四章 結論與建議

為建構土石流災害智慧防災系統並提升災害預測的準確性，本研究以巨量資料的概念為基礎，將所有土石流觀測站的監測資料皆一同納入分析，透過不同監測因子與土石流發生事件之探討，釐清各種監測因子用於預測土石流發生之可行性。接著利用詞袋模型(Bag-of-Words)，將時序性資料重新解讀，初步於神木村觀測站測試後，發現該模型更能表達一段時間內監測值連續行為的特徵，所發布之預警的準確度較原始時序性降雨監測資料高。因此，本研究便以詞袋模型為基礎，加入過濾門檻，並且建立自動化預警之控制流程，發展土石流災害警戒系統。

本研究以神木、豐丘、上安及郡坑等四個測站進行試驗，並調整出最合適各測站的參數，且將警戒以訓練年的七日平均土壤含水量及累積降雨量作為過濾門檻，使警報的發布更為嚴謹。在神木與豐丘兩測站的命中率及可信度的表現，皆較現行的土石流預警方式為佳；上安和郡坑站在歷史上無土石流災害發生，詞袋模型的誤報次數亦較現行方式為少。因此，未來或可以此模式為基礎，發展土石流預警系統，來提升預測土石流發生之準確性。

然而以神木村的測試結果而言，仍然有兩場真實土石流事件之漏報。透過漏報事件的累積雨量檢視後發現，該兩場事件的前七日累積雨量僅約 56.5 及 85.0mm。此結果與土石流發生要件的「充足的水量」顯然並未完全相符。而造成此情況的因素，是否由於雨量觀測站設置於土石流的流動段，而較無法反映實際土石流溪流集水區的整體雨量所致。因此，在雨量站設置位置的選擇，或是降雨空間分布的推估，應是未來可延續探討的研究方向。

參考文獻

1. Caine, N., "The Rainfall Intensity Duration Control of Shallow Landslides and Debris Flows." *Geografiska Annaler* Vol.62, 23-27, 1980.
2. EJ Candès, X Li, Y Ma, and J Wright. "Robust principal component analysis?" *Journal of the ACM*, 2011.
3. Fan, Jen-Chen, et al., "Effects of landslide and other physiographic factors on the occurrence probability of debris flows in central Taiwan." *Environmental Earth Sciences*: 1-17, 2015.
4. J. Lin and Y. Li. "Finding structural similarity in time series data using bag-of- patterns representation," In *International Conference on Scientific and Statistical Database Management*, 2009.
5. J. Dean and S. Ghemawat. Mapreduce: simplified data processing on large clusters. *Communications of the ACM*, 51(1):107--113, 2008.
6. Keefer, D.K., R.C.Wilson, R.K. Mark, E.E. Brabb, W.M.Brown, S.D. Ellen, E.L. Harp, G.F.Wieczorek, C.S. Alger, and R.S Zatzkin., "Real-Time Landslide Warning During Heavy Rainfall", *Science*, Vol.238, 921-925, 1987.
7. Takahashi, T., "Debris Flow. Disaster Prevention Research institute." *Kyoto University, Kyoto, Japan*, 1981.
8. Wieczorek, G. F., "Effect of rainfall intensity and duration on the debris flows in central Santa Cruz Mountains." *California, Geological Society of America Reviews in Engineering Geology*, 7, 93-104, 1987.
9. Y.-J. Lee, Y.-R. Yeh, and Y.-C. Wang, "Anomaly Detection via Online Over-Sampling Principal Component Analysis," *IEEE Transactions on Knowledge and Data Engineering*, volume 25, number 7, pages 1460-1470, July 2013.

10. 李明熹，“土石流發生降雨警戒分析及其應用”，國立成功大學水利及海洋工程研究所博士論文，2006。
11. 林炳森，“土石流監測與預報系統之研究(I)”，NSC87-2621-P-005-010-，行政院國家科學委員會，台北，台灣，1999。
12. 范正成、姚正松，“台灣東部地區土石流發生的水文及地文條件應用於土石流預警之初步研究”，中華民國第一屆土石流研討會，125-139，1997。
13. 范正成、吳明峰、彭光宗，“豐丘土石流發生臨界降雨線之研究”，地工技術，74：39-46，1999。
14. 范正成、吳明峰，“一級溪流土石流危險因子及其與臨界降雨線之關係”，中華水土保持學報，32(3)：44-53，2001。
15. 張守陽、黃榮堂、李璟芳，“機械視覺辨識土石流發生之初步研究”，第十一屆水利工程研討會論文集，台北，台灣，第 K5-K10 頁，2000。
16. 黃效禹，“地文因子對土石流發生機率之影響研究”，國立臺灣大學生物環境系統工程所博士論文，2015。
17. 黃清哲、葉智惠、尹孝元、王晉倫，“地聲探測器應用於土石流監測之研究”，第四屆海峽兩岸山地災害與環境保育學術研討會，pp. 49-56，2004。
18. 詹錢登，土石流危險度之評估與預測，中華水土保持學報，25(2)：95-102，1994。
19. 詹錢登，“土石流降雨警戒模式及警戒基準值之改善研究”，水土保持科技與研發展望研討會，4-1~4-15，2003。

20. 劉格非，李欣輯，“地聲探測器之應用”，第二屆海峽兩岸山地災害與環境保育研究，9月，台中，台灣，pp.161-170，2000。
21. 劉哲欣，“利用降雨及地下水水質資料進行土石流發生之監測及分析”，國立臺灣大學生物環境系統工程所博士論文，2009。
22. 謝正倫、陳禮仁，“土石流潛在溪流之危險度評估方法”，中華水土保持學報，24（1）：13-19，1993。