

# 開發機器學習之情資分析技術與應用

## Development of machine learning technology for disaster intelligence analysis and application

主管單位：國家災害防救科技中心

張子瑩 劉致灝 郭政君 張智昌 陳俊元

### 摘要

災防科技中心自 99 年起，建構災害應變決策輔助系統，整合多元災防資料，包括基礎、監觀測、模式、以及災情資料等，超過 20 單位、近 300 項圖資，目的是建構一套仍讓中央與地方防災人員，具共同圖像為目標的防災系統。在多來源的災防資料中，透過機器學習的方式，能協助人員偵測重要訊息，或進行異常值偵測，以建立「智慧化」的第一步。本年度在機器學習上，著重在影像重要訊息的偵測，透過卷積神經網路法，從災情照片的特殊地標，再進一步判斷招牌上的電話資訊，得以準確進行電話與地址座標配對，以協助災情點位確認。另外也透過關聯規則探索技術，分析災防示警資料間的關聯性，以協助偵測可能因傳送系統設定、人為操作錯誤等，而發生短時間重複傳送、發布相同內容、疑似感測器(Sensor)故障造成短時間不斷發布/解除訊息，或未發布的示警即發布解除訊號等狀況，以輔助人員判斷資料的正確性，提高傳遞示警訊息的品質。

**關鍵詞：**機器學習、自動分類、自動判釋

### Abstract

Since 1999, the Disaster Prevention Science and Technology Center has established a disaster decision support system that integrates multiple relevant data, including basic maps, monitoring and observation data, numerical data, and disaster reports, etc.. Now the system has been integrated more than 20 units and nearly 300 layers in it, and a common operational picture for central and local governments has constructed successfully. Using machine learning helps operators to detect important information or anomaly from the multi-sources of disaster prevention data is the first step to implement "smart system". In 2019, we focused on the detection of important information in the image via machine learning, and the convolutional neural network method was used in this study. We can detect the special landmarks of the disaster photos, and then automatically classified the phone numbers on the signboard. Finally the search engine helped to find the accurate address to give a coordinate information. It could be

accurately matched and confirmed the disaster points. In addition, we also uses association rule exploration technology to analyze the correlation between different kinds of warning data. The automatic interpretation function used to assist in the detection of the same content in the short-term repeated transmissions, suspected sensors failure causing continuous release / dismissal of messages for a short period of time, or releasing a error signal without a warning due to transmission system settings, human operation errors, etc.. It could be used to help operators to interpret the correction of the data and furthermore, to improve the quality of transmitting warning messages.

**Keywords : machine learning, automatic classification, automatic interpretation**

## 一、前言

防災科技中心依災防資訊研判的需求，蒐集來自不同類型的數據資料，包括具連續性的感測站資料，如水位站資料，或是因災防特性的非連續性檔案資料或影像資料，如示警資料或災情照片等。大量的災防數據處理上，多仰賴電腦程式自動化的資料處理、派送及傳遞。連續性的感測資料，較容易透過歷史的統計值，進行自動化資料前處理。但對於不連續性的資料，如防災示警資料，及災情照片等，為能提升資料傳遞品質，始下游端應用端能有更穩定的服務開發，因此在資料的分類及警示功能上，則希望能有精進。

## 二、研究分析及成果

### 2.1 開發機器學習自動化分類

隨著資料量的逐年成長，以及機器學習的技術成熟，運用卷積神經網路 (Convolutional Neural Networks, CNN) 進行災情照片訓練及判釋，透過1000張招牌影像作為訓練樣本及模型建置，對於重要地標招牌的特徵，可有效辨識，採用隨機選取的200張影像進行連鎖商店招牌(統一超商、全家便利商店、全聯福利中心以及大潤發量販店)特徵辨識的驗證，正確率可達95%。因此再以相同的學習模式，針對招牌上的數字進行特徵值學習，以電話號碼辨識為主，透過電話號碼資訊，可以容易搜尋到正確地地址位置，以獲得更準確地位置定位座標。

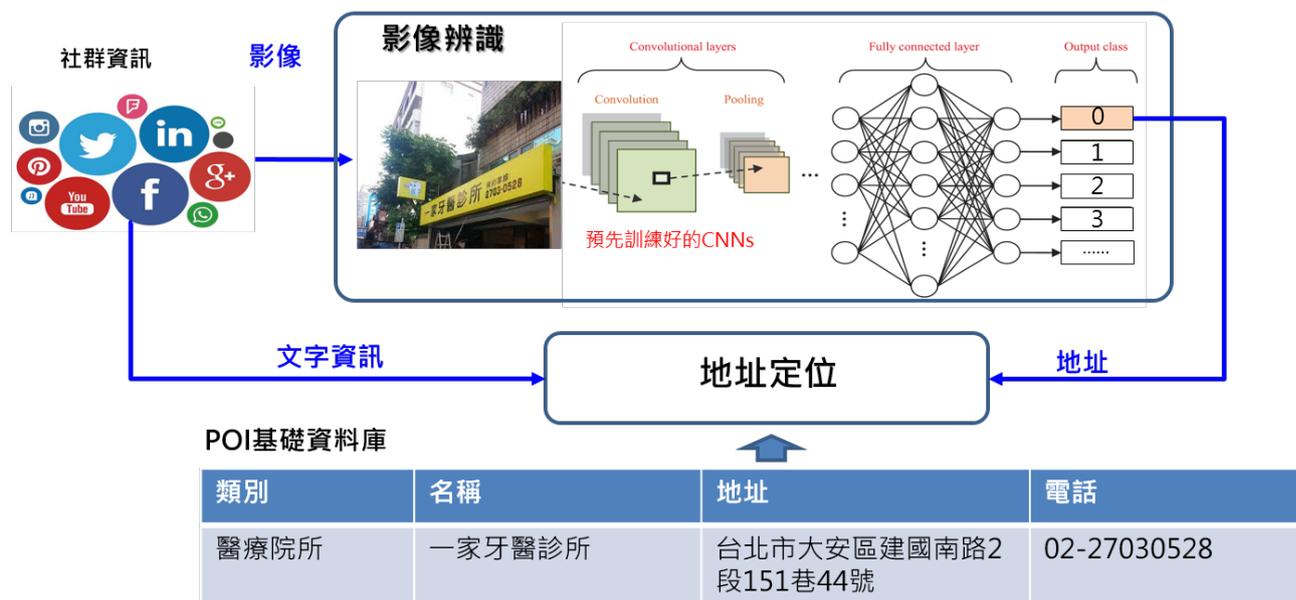


圖 1 招牌電話號碼定位案例

### 2.2 機器學習自動資料品質告警

防救災示警及民生示警，已成為對民眾端，最直接且最重要的災防資訊，但示警的資料可能由於提供部會端的系統設定問題，出現短時間重複傳送或是傳送說明不清的資訊，導致自動化系統在傳送給民眾服務時，造成品質不穩定的問題。由於示警非連續性資料，對資訊系統而言，就像是突然的事件觸發後，再產生一連串的程序執行，但因為是突發性的資料特性，是否能解決完全不可預期狀況，提供管理者的檢核參考。以2018年8月22日至24日南部豪雨為例，依序發布了降雨示警、道路封閉示警、土石流示警、停班停課示警、水庫放流示警、市話通訊中斷示警和河床水位高示警等，以顯示示警具有地域及時序特性，因此本年度採用降雨示警與相關示警進行關聯性分析進行學習，以進一步分析示警的發布特徵。

108年度已整理2015年至2018年的示警資料，透過關聯規則探索（Association Rule Discovery）方式，也就是機器學習中非監督式學習的一個演算法，找出資料共同出現的情況，分析資料間的相似性與相關性，透過分析出的每一個規則，以支持度(Support)、可信度(Confidence)、增益度(Lift)三種規則特徵來找示警共同出現的機率。透過Apriori 演算法，在可信度(confidence)為0.2與支持度(support)為0.05的條件下，先找出112 條規則，取出以降雨示警為首的示警的規則共52條，再把重複的規則去除後，最後獲得21條規則如表1。其中21條規格的三種特徵的分布情形，雖然支持度的機率都不高，僅第18條規則高於0.5，但可信度大於0.5且增益度皆有較高的趨勢，可為後續當相似的預警發布時的參考。以第2條規則說明，當降雨示警、水庫放流示警與雷雨警報發生時，同時會伴隨發布淹水示警的可信度(條件機率)為0.735，其增益度為2.047，代表當看到降雨示警、水庫放流示警與雷雨示警一起發布時，有很大的可能性會伴隨發布淹水示警，且遠大於降雨示警後發布淹水示警。

表 1 關聯性規則

編號	A	B	支持度	可信度	增益度
1	降雨示警, 淹水示警, 道路封閉示警	河川高水位示警	0.063	0.331	2.103
2	降雨示警, 水庫放流示警, 雷雨示警	淹水示警	0.091	0.735	2.047
3	降雨示警, 水庫放流示警, 道路封閉示警	河川高水位示警	0.074	0.311	1.974
4	降雨示警, 雷雨示警	淹水示警	0.123	0.615	1.712
5	降雨示警, 河川高水位示警, 水庫放流示警	淹水示警	0.061	0.603	1.679
6	降雨示警, 河川高水位示警	道路封閉示警	0.121	0.774	1.594
7	降雨示警, 水庫放流示警, 道路封閉示警	淹水示警	0.128	0.534	1.488
8	降雨示警, 河川高水位示警	淹水示警	0.079	0.500	1.392
9	降雨示警, 淹水示警	水庫放流示警	0.241	0.674	1.331

10	降雨示警,水庫放流示警,道路封閉示警	雷雨示警	0.061	0.255	1.271
11	降雨示警,河川高水位示警	水庫放流示警	0.101	0.642	1.268
12	降雨示警,雷雨示警	水庫放流示警	0.123	0.615	1.215
13	降雨示警,地震報告	道路封閉示警	0.065	0.587	1.209
14	降雨示警,淹水示警	道路封閉示警	0.192	0.537	1.107
15	降雨示警,雷雨示警	道路封閉示警	0.104	0.519	1.069
16	降雨示警,強風示警	道路封閉示警	0.050	0.500	1.031
17	降雨示警,地震報告	水庫放流示警	0.056	0.507	1.001
18	降雨示警	水庫放流示警	0.506	0.506	1.000
19	降雨示警	道路封閉示警	0.485	0.485	1.000
20	降雨示警	淹水示警	0.359	0.359	1.000
21	降雨示警	雷雨示警	0.200	0.200	1.0000

108 年度透過關聯規則技術，分析五縣市中五大示警與降雨示警之歷史樣態機率，優先鎖定降雨示警發布數量最多的五大縣市，包含新北市、宜蘭縣、屏東縣、基隆市與高雄市，以降雨示警為首搭配最可能伴隨降雨示警之五項示警，包含淹水示警、河川高水位示警、土石流示警、水庫放流示警與道路封閉示警，計算縣市示警發佈機率。其中降雨示警分為大雨(輕微)、豪雨(中等)、大豪雨(嚴重)、超大豪雨(極嚴正)四個等級，圖 2 為新北市為例展示不同月份為單位，不同降雨等級示警計算出與其他相關示警發佈之機率。



圖 2 新北市發布示警機率圖

依新北市的分析可得知，於降雨在各示警在不同月份發布的機率有明顯的不同，新北市土石流示警集中於 6-10 月，不同嚴重程度的降雨示警發佈條件下，發生五項相關示警發佈的機率也不同。因此依據分析數據，建立自動資料品質告警透過模組，系統管理者可設定示警資料品質自動告警，如當系統接收到新北市於 2019 年 1 月份發布土石流示警時，透過模型計算新北市 1 月份會發布土石流示警的機率接近 0，再配合獲得是否有豪雨等級的降雨示警發布時，此模組就會對管理者提出告警訊息，合理懷疑此則土石流警戒為異常示警，需管理者介入進行資料釐清後始得發送。

### 2.3 開發微服務架構提升對外服務

由於資訊基礎服務資源有限，如單一機房的擁擠或是機房租賃昂貴，本年度透過試作微服務架構，依不同機房的資源設定不同的部署數量，配合防災系統以資料即時傳遞的特性，評估後續架構改變的效益，以確保防災系統的高可用性。

#### (1) 微服務架構網站分析評估

傳統對外服務系統往往採用單體式應用程式(Monolithic application)設計，將所有的功能撰寫於單一系統內，相對於傳統單體式服務的概念，微服務的概念是將一個複雜且龐大的單體式應用程式系統(Monolithic)拆分為多個具有不同功能的小服

務(Microservices，又稱為微服務)。採用微服務架構的應用程式開發的優點是：

- A. 開發人員僅需專注於小範圍程式：每個服務僅負責單一功能，當功能有增修時開發人員僅需專注小範圍程式的增修。
- B. 更新速度加快：當功能有增修時，僅需更新相對應的服務，不須完整更新所有服務系統。
- C. 不須仰賴單一技術：單體式應用程式必須選擇單一程式語言開發，而在微服務架構應用程式則不受限制。開發人員可以針對服務挑選最合適的開發語言進行開發，可以不同程式語言建構服務系統。
- D. 有效利用資源：可依各項服務所需資源不同而部署不同數量的服務(如圖 3)。

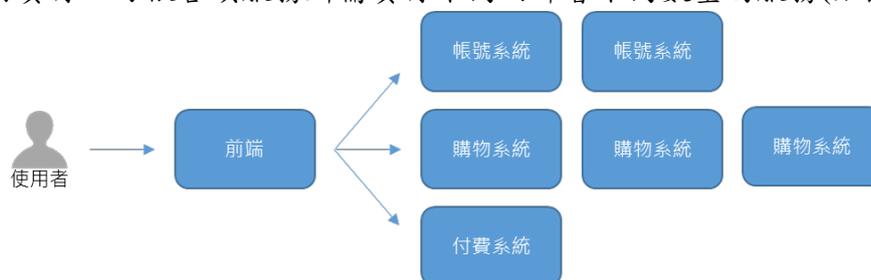


圖 3 微服務的擴展性

採用微服務架構的缺點則是：

- A. 開發的複雜性：單體式應用程式架構已經發展多年，許多成熟技術以廣為開發人員熟知。微服務架構不同於單體式應用程式思維，在系統設計階段之際，就必須規劃將服務拆分成許多細小的服務，並建立起各服務間的通訊與關係，使得微服務架構的開發複雜性相對於單體式應用程式高的許多。
- B. 偵錯相對耗時：隨著服務部署數量的增加，將會增加偵錯的複雜度。當有錯誤發生時，必須額外找出資訊流的流向，若同時與多個子服務通訊，則亦須同時對多個子服務查詢錯誤原因，造成維護時間成本的增加。

## (2) 微服務架構網站試作-在混合雲上部署圖磚服務

圖磚服務常用於防災相關的展示系統中，用以展示空間化的地圖資訊或影像資訊，因此本次開發的微服務平台以圖磚服務為底圖。本次部署的圖磚系統微服務架構，於台北東七機房、台中文心機房兩地的私有雲以及一地的 Google Cloud Platform(GCP)雲端伺服器等三地共佈設 9 台伺服器，每個機房可依資源考量(如網路頻寬、公有雲平台費用)分別設定不同服務部署數量，其中東七機房部屬 4 台伺

服务器，文心機房部屬 3 台伺服器，GCP 雲端機房部屬 2 台伺服器，透過程式碼管理平台，可顯示圖磚系統在不同機房及伺服器上，持續性發佈的服務的情況，圖 4 畫面上成功展示不同圖磚，是來自不同機房的服務，其中 Server: X-Y 表示機房 X 內的第 Y 個服務，畫面上 Server: 1-3 則表示此服務來自於東七機房內的第三台伺服器所提供的圖磚服務。

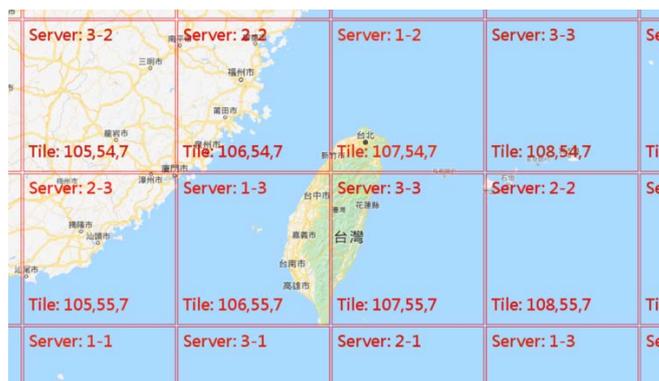


圖 4 圖磚系統之系統畫面

此外，考量網站系統以微服務架構不屬於公有雲及私有雲上，故以導入 DevOps(Development 和 Operations 的組合詞)技術，以提升網站系統自動部署、快速交付的速率。在 DevOps 的技術操作下，可使開發團隊與營運團隊更加緊密的合作，因此在開發、測試、部署、擴展、功能增修以及資訊安全等多工的情況下，透過微服務工具展現高度自動化，表 2 分析網站或系統開發，微服務配合 DevOps 的優勢，可大幅縮短上線時間，以能同時因應需求及保持系統穩定性與可靠性。

表 2 DevOps 與傳統方法的比較

	傳統	導入 DevOps
開發週期	以周或月為開發週期	以分鐘或小時為開發週期
源碼檢測	常須需跨單位進行 約 1~3 日	使用工具自動檢測 約 1~10 分鐘
測試	人工測試 約 0.5 日	使用工具自動檢測 約 1~10 分鐘
部署	人工部署 約 1~7 日	自動部署 如有異常可隨時退回上一版本 約 10~30 分鐘
資安弱掃	外部弱掃機構進行檢測 約 2~10 日	內部弱掃工具定期自動檢測 約 1~10 分鐘
監控	花費大量人力進行監控及	由工具收集及分析日誌，過濾出主要異常

	處理日誌	事件，並自動觸發對應機制。如可在流量大時自動增加雲端上的服務。 人工僅需處理意外事件
成本	開發、部署、維運成本都很大	前期導入微服務成本較大， 後期開發成本下降，部署維運成本亦低於傳統開發

### 三、結論與建議

- 機器學習自動化分類功能：  
透過機器學習建立模型，建立災情影像類資料自動分類模組，可自動偵測災情影像內招牌樣式及招牌內電話，以自動化搜尋定位。
- 機器學習自動示警功能：  
對於跨部會示警資料進行盤點，以關聯性分析法建立不同縣市及月分之示警出現機率模型，將異常資料提前或即時偵測出，並提供管理者告警之功能。
- 微服務架構評估：  
完成微服務架構網站分析，評估傳統單體式應用程式架構及微服務架構之優缺點及試作圖磚微服務架構之應用案例。