

淺談大數據技術於森林經營應用

謝漢欽¹、黃俊元²

乘著以大數據為基礎的人工智慧發展浪頭，林業研究趨向智慧林業(smart forestry)發展之聲不絕於耳。本文試圖從相關研究文獻探討傳統森林經營所需的森林監測下的林業資訊系統，在面對迅速、大量獲取且多元的林業大數據(forestry big data)時、如何藉機改變原來的樣貌，採用資料科學方法，著重在林業大數據系統架構下，從資料獲取、儲存、查詢、分析及應用層面的新技術介紹；特別指出大數據的不同儲存方式與快速計算能力的需求，最後介紹3個國外研究的應用實例，供台灣發展智慧林業研究的參考。

大數據與智慧林業

以傳統的森林監測方法將資料記錄於結構化的關聯表單有增加外，過去數十年來由於網際網路、智慧物聯網(AI internet of things)及移動互聯網(例如手機等)的蓬勃發展，由其所產生的半結構化與非結構化資料量以指數倍增。為了從此巨量資料中快速地以資料挖掘(data mining)方法掘出有用的規則、模式及提供決策資訊；如果以傳統統計分析方法為基礎的電腦計算速度與資訊系統的儲存與分析能力常無法勝任，此時大數據的技術因運而生。大數據的技術除了必須具有獲取和儲存多元資料格式的能力外，更著重在其資料處理技術與處理工具的整合應用層面上。

大數據的主要特徵在所謂的3V，即「大的資料量(volume)」，一般指資料量

至少高出TB=1,024GB以上，「資料格式多樣性(variety)」，以及「資料產生速度快(velocity)」；後來加上「資料的品質與可信度(veracity)」，以及是否「具有高度(商業)價值(value)」備受重視且不容忽視，形成大數據的5個重要的V。目前大數據主要廣泛應用於都市基礎設施的建構、智慧醫療、智慧格網(smart grid)、智慧交通、智慧防救災網路、社會服務網路(social service networks)，以及農業和林業監測上。

有關智慧林業的觀念目前定義尚不明確，就技術而言，主要是以過去數位化林業，也有人稱其為精準林業(precision forestry)的技術，例如：遙感探測(RS)、地理資訊系統(GIS)及全球衛星定位系統(GPS)，3S空間技術整合於森林資源調查與林業經營資訊決策支援系統等技術為基礎，持續發展而來。

智慧林業另外增加了大數據技術，如分散式雲端儲存與平行運算(cloud computing)、智慧物聯網(強調邊緣運算的應用)、移動式網際網路(mobile internet)、資料挖掘、資料科學(data science)分析法(主要是指在人工智慧領域下之統計分析、模式判別、機器學習與從機器學習分支的深度學習演算、辨識或預測模式建立等方法)，有關人工智慧領域的發展請參考圖1，以及其他新一代資訊科技的整合應用，也可稱之為新一波的數位轉型。

透過智慧林業可以兼顧森林資源的經營與森林生態系統的健全，使得森林經營更

^{1,2} 林業試驗所·森林經營組、太麻里研究中心

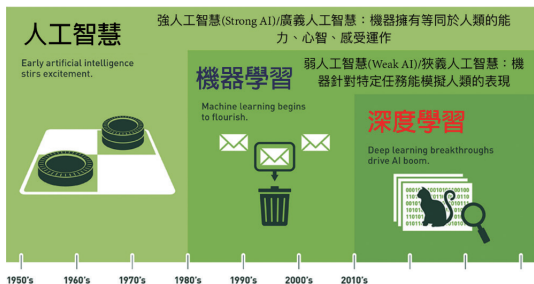


圖1 人工智慧領域的發展與現階段以資料為基礎的機器學習演算法。(NVIDIA Blog, 2016)。

數位化、有感知、相互連結及有智能表現。其中特別強調擁有林業大數據是智慧林業系統的基礎，透過以密集資料為燃料的驅動典範，有利於林業大數據的統計分析與資料挖掘、機器學習等演算建模，並以視覺化呈現互動式的決策分析結果，據以輔助森林經營決策；目前應用大數據技術於林業領域仍處於剛起步的研究發展階段。

森林經營大數據的形成

在長期森林經營的過程，必須以連續森林監測為前提，應用森林取樣調查技術，蒐集有關森林資源的分布、組成、結構及外部的干擾(disturbance)等隨時間變化的動態資訊，並透過各種政府開放資料(open data)以GIS進行資料整合與分析，進行森林經營規劃情境分析，以提供最佳的經營決策與施業方案。因此能獲取上述各種格式的森林經營大數據，

以精確可靠的資料作為驅動，達成有效預測、診斷模式的建立、驗證及部署，據以輔助經營的決策與行動，方能有效維護森林生態系統，達成永續森林經營目標。

從2011年大數據時代來臨，森林調查工作也面臨勞力缺乏且投入成本高的年代，此時森林監測主要透過地面監測儀器與遙感探測技術獲取大量資料，前者包含了GPS、各種用於定點或移動式(手持式、車載)感測器、地面光達掃描儀(laser scanner)，以及近景攝影量測。後者則包括了無人機系統(UAV)攝影或空載光達(LiDAR)所獲取的3D點雲(cloud point)及影像，以及各種不同光譜、空間及時間解析度的航測及衛星影像等。以地面為基礎的感測儀器可組成無線感測器監測網(WSN)，蒐集區位範圍的即時監測資料，可與從航遙測攝取之大範圍、多尺度影像資料互補運用。森林監測透過新的技術可提高大量資料

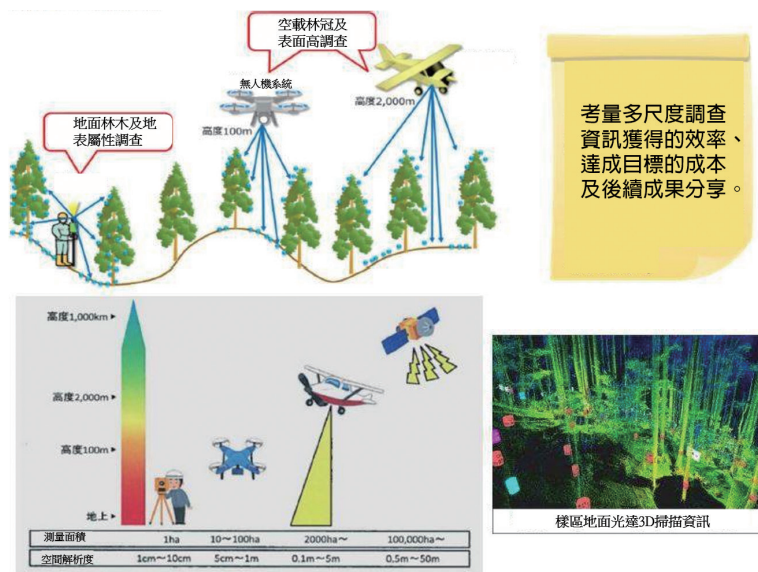


圖2 日本智慧林業運用新技術進行森林調查與監測示意圖。(日本林野廳, 2019)

獲取的效率並增進資料的精確度，圖2為日本智慧林業的森林調查與監測示意圖。

以上述新技術獲取的資料，將其建立在傳統的森林資訊系統或森林經營決策地理資訊系統的關聯式資料庫系統中，進行各項目的資料存取、查詢、分析與資訊視覺化呈現，會面臨大數據特徵前3個V的問題，特別是面對多維度、多樣化空間大數據時會導致資料處理的困難。尤其是隨著各種感測器、衛星及無人機在大面積範圍內所獲取大量高光譜(hyper-spectrum)、多時期、高空間解析度影像，使得監測資料量急遽暴增，傳統的資料庫系統與資料分析的電腦計算能力已無法勝任；尤其在面臨全時間森林災害監測與即時預警，需將所獲的大量資料進行快速計算，常無法突破瓶頸。因此，需要以大數據專有的技術來解決這些問題，也就順勢創造了林業經營監測技術轉型的新契機。

林業大數據技術簡介

數位化林業時代面對低解析度、小範圍的遙測資料及有限的地理資訊圖層，進行資料處理與分析時，使用單一電腦便足以勝任。但到了大數據時代的林業資訊系統，不只是能容納快速獲取的資料龐大，其主要的挑戰在於資料組成的複雜程度及多樣的資料結構。傳統單機運算的林業資訊系統，無論在資料庫儲存型態、資料查詢、分析、挖掘、建模所需的快速計算與即時提供決策方案等，往往無法到位。然而林業大數據處理技術如果能善用處理大數據之硬體與軟體系統組合計算平台，可以解決上述問題。

近年來與林業大數據相關的大數據處理

平台研究發展迅速，如: Hadoop、Spark、及Strom等，各有所長的開放原始碼(open source)的大數據儲存及處理平台，足以解決大數據中需要的大多數非結構化資料格式、分散資料雲端儲存、多機連結雲端平行計算、及多樣化資料查詢、分析、資料挖掘、機器學習演算與建立應用模式的功能。因此，一個理想的林業大數據管理組合平台，應具有從資料獲取、儲存、查詢、分析到應用5個層次的架構，如圖3所示。資料獲取層必需能蒐集不同格式的林業資料，以森林經營大數據而言，雖然必須面對各種型態結構複雜的資料格式，但在發展初期的大數據，主要著重在遙測的網格式資料及地理資訊圖層的向量式資料。儲存層則著重於資料如何合理的儲存於可持續擴增、具有容錯機制的分散資料叢集(clusters)，以及各類資料集之間的指標(index)結構的彈性設計。查詢層則聚焦在需求條件下資料的讀取與過濾時，必需符合平台記憶體體的讀取高效率與

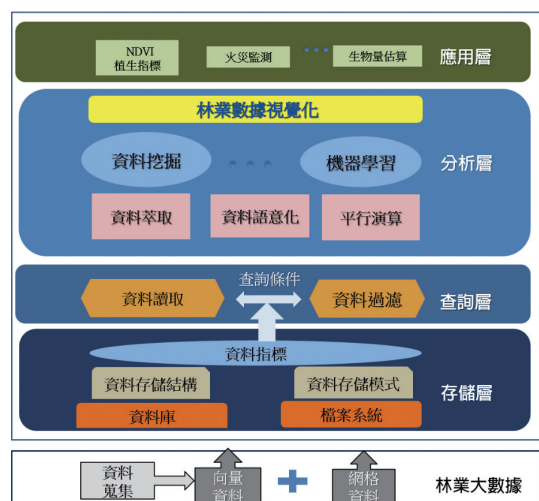


圖3 理想的林業大數據管理平台具有5個層次架構。(Zou et al. 2019)

最小成本原則。分析層則著重在資料科學的高階統計分析、資料挖掘、機器學習演算(常需多層次遞迴計算，需耗用平台大量的計算資源)，及分析結果的交談式動態視覺化的呈現。應用層可將傳統林業資訊只以檔案格式或關聯式資料存取與處理方式，轉換到兼具分散式檔案系統、非關聯式資料庫、以物件導向分析之大數據處理平台，才能享有對林業大數據高效能資料處理與分析能力。此一林業大數據管理平台架構可部署在單一電腦集群或與具有結構化設計的雲端架構。

森林經營大數據應用實例

本文簡要舉出3個應用大數據技術的國外研究案例：

1. 森林植生分類與變遷分析：目前國際上在大尺度森林覆蓋範圍，結合了多尺度光學遙測(包含UAV)之多時期影像，特別著重在使用高光譜影像與由空載光達資料產生的數值地形(DEM)與數值表面高(DSM)，高解析度影像的質地(texture)特徵，以物件為基礎(object based)的影像分類程序，進行森林類型或樹種的分類。由於分類程序多半已使用到機器學習演算與深度學習(deep learning)等多重遞迴計算，因此需要用到大數據處理平台的計算資源及其所提供可平行計算的各種分類演算法。有了高準確度森林植生分類模式的建立，再結合多時期的遙測影像就可進行森林變遷分析，並經由長期連續性偵測與大數據處理平台技術的支援，能提供即時的變遷資訊。
2. 森林生長參數的推算：森林生長參數包括了林木的胸徑、樹高、葉面積指數、樹冠

層疏密度等。這些參數在不同時空尺度需求上，可整合雷達、空載光達、高光譜及UAV等遙測資料及其衍生的植生指標(例如NDVI)，連結地面逢機選取的樣區調查資料，經由多層次遞迴演算後建立可靠的推估模式。當面臨林地覆蓋面積龐大時，更需要借助大數據處理平台的分散儲存與平行演算能力的奧援。

3. 森林災害即時監測：森林災害的監測包括森林火災預防及病蟲害的偵測，森林火災災害的監測需要多時期、高光譜解析度及低空間解析度衛星影像(如MODIS衛星影像)，一般還要連結由地面設置的無線感測器網路的部署，獲取相關的監測資訊，利用大數據平台的平行演算技術，建立林火預測模式與及早期警告發布系統。森林的病蟲害監測則需要結合高光譜、高時空解析度的遙測資料，進行長期持續的監測，主要是利用不同時期的高光譜變異資訊進行影像分類及辨識偵測，同時也需要結合地面感測器的機械視覺技術(machine vision technology)，以達成即時的偵測與病害發布，因此也要借用大數據處理平台的雲端運算技術來達成。⊗

(參考文獻請逕洽作者謝漢欽，email: mickey@tfri.gov.tw)