

# 機器學習與天文學：無人大型巡天計畫



李見修／中央大學天文碩士，慕尼黑天文博士。  
目前在日本昴宿天文台擔任支援天文學家。

## 天文學：開放的巨量資料

提到機器學習在天文領域的應用，大家第一個想到的便是星體分類。這是因為現代的大型巡天計畫已累積了相當可觀的巨量資料。而在計畫的主要科學課題之外，這些資料還蘊藏無限可能，但可惜的是通常計畫成員已無法兼顧其他課題。與其他學科不同的是，天文學家們知道蒼穹之大，非一己之力所能參透。與其將資料鎖著不見天日，不如開放給全世界（包括普羅大眾）一同使用。恆星演化、伽瑪射線爆、以及重力透鏡的先驅波登。波蘭天文學家普欽斯基（Bohdan Paczyński）就曾說過這麼一句話：「只有將資料公開，與對手共享，才能讓我們的觀測數據發揮最大的效益。」

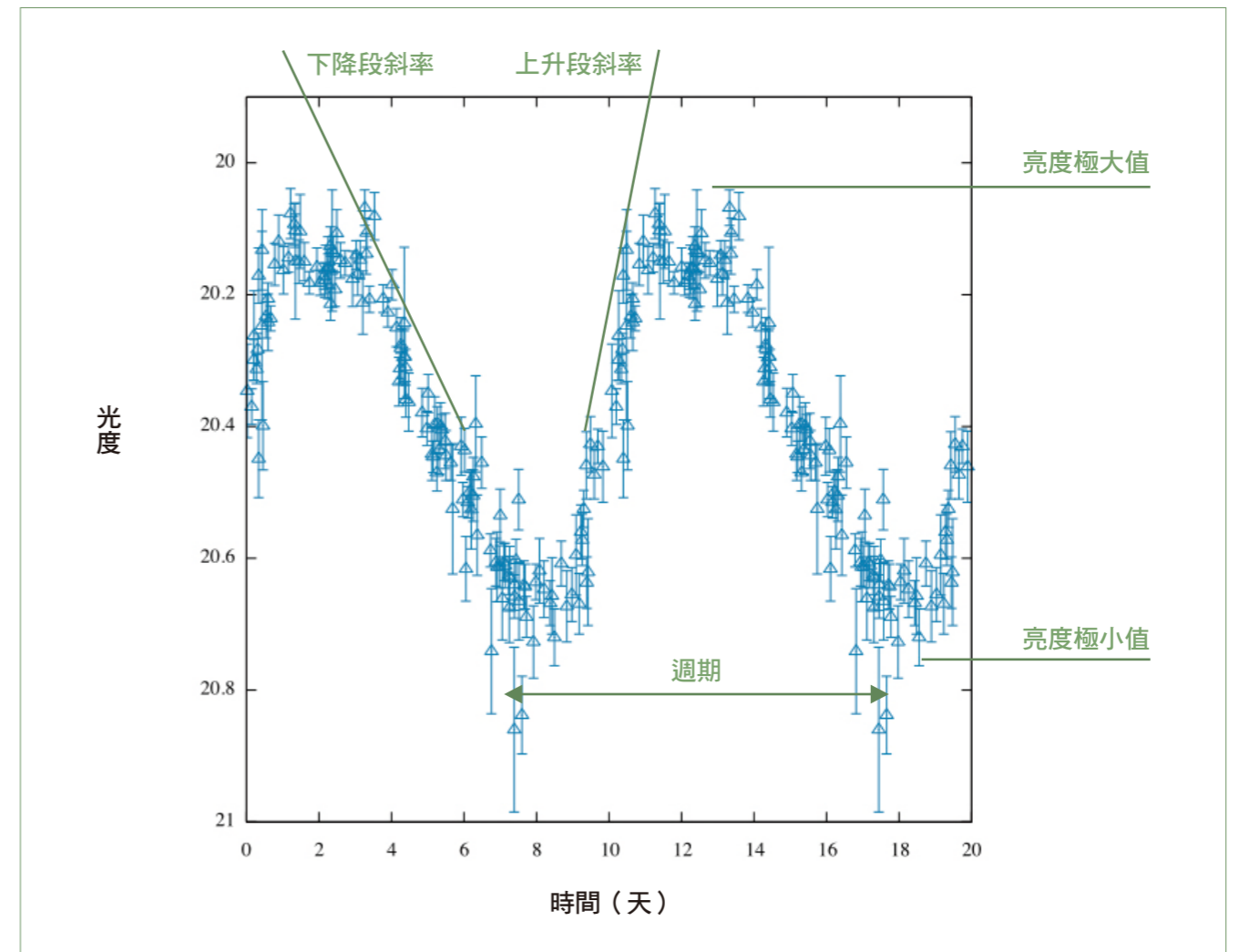
## 使用機器學習代替目視分類

但是原始資料整理起來相當繁瑣又曠日廢時，於是天文學家想到利用機器學習來分析，一方面可以減少投入的人力，另一方面也能建立起不同的目錄，讓有興趣的天文學家自行取用。舉例來說，像是「全天自動巡天計畫（All Sky Automated Survey, ASAS）」，利用兩具口徑 15 公分的望遠鏡，分別從夏威夷以及智利觀測全天。這個計畫自 1997 年開始、平均每兩天巡天一次，持續至今日仍在進行，對超過兩千萬顆亮星留下長時間的觀測資料，可

供變星分析使用。然而巨大的資料量，加上變星種類繁多，要一一目視分類實屬不易。加州大學柏克萊分校的約瑟夫理查斯等人因此想到使用機器學習的方法來分析這巨量的資料，透過擷取光變曲線的特徵，來依此分類不同的變星（圖一）。

機器學習同時也可以運用在星系分類上，並依此取得星系的基本性質。舉例來說，由史隆基金會所贊助的史隆數位巡天計畫（Sloan Digital Sky Survey, SDSS，臺灣中研院天文所亦參與其中）利用位於美國新墨西哥州的一架口徑 2.5 米的望遠鏡，對北半球的天空進行可見光多波段（grizy 濾鏡）的巡天，留下了深且廣的大範圍北半球天區影像，可供天文學家們研究星系在不同環境、不同宇宙年齡時的性質，讓我們進一步了解星系的演化。但在使用這筆資料時，首要之務便是能對數以萬計的影像分類。可是即使是星系專家們，要看完所有的影像也是需要許多時間。比利時根特大學的桑德帝勒曼等人因此想到使用類神經網路的方式分類星系。他們的方法除了將史隆數位巡天計畫資料裡的星系，最基本的分類為橢圓星系與螺旋星系外，更進一步細分螺旋星系是否有棒狀結構，以及螺旋星系是屬於正視（face-on）亦或是側視（edge-on）等。而在星系分類好後，便需決定星系的距離。這是因為距離我們越遠的星系越古老，因此根據不同的距離，我們可以推測星系在宇宙不同年齡時的演化性質。

一般來說，要測定星系的距離，最準確的方法是透過拍攝星系的光譜。因為宇宙膨脹的關係，當星系離我們越遠，它遠離我們的速度也越快。根據都卜勒效應，當一



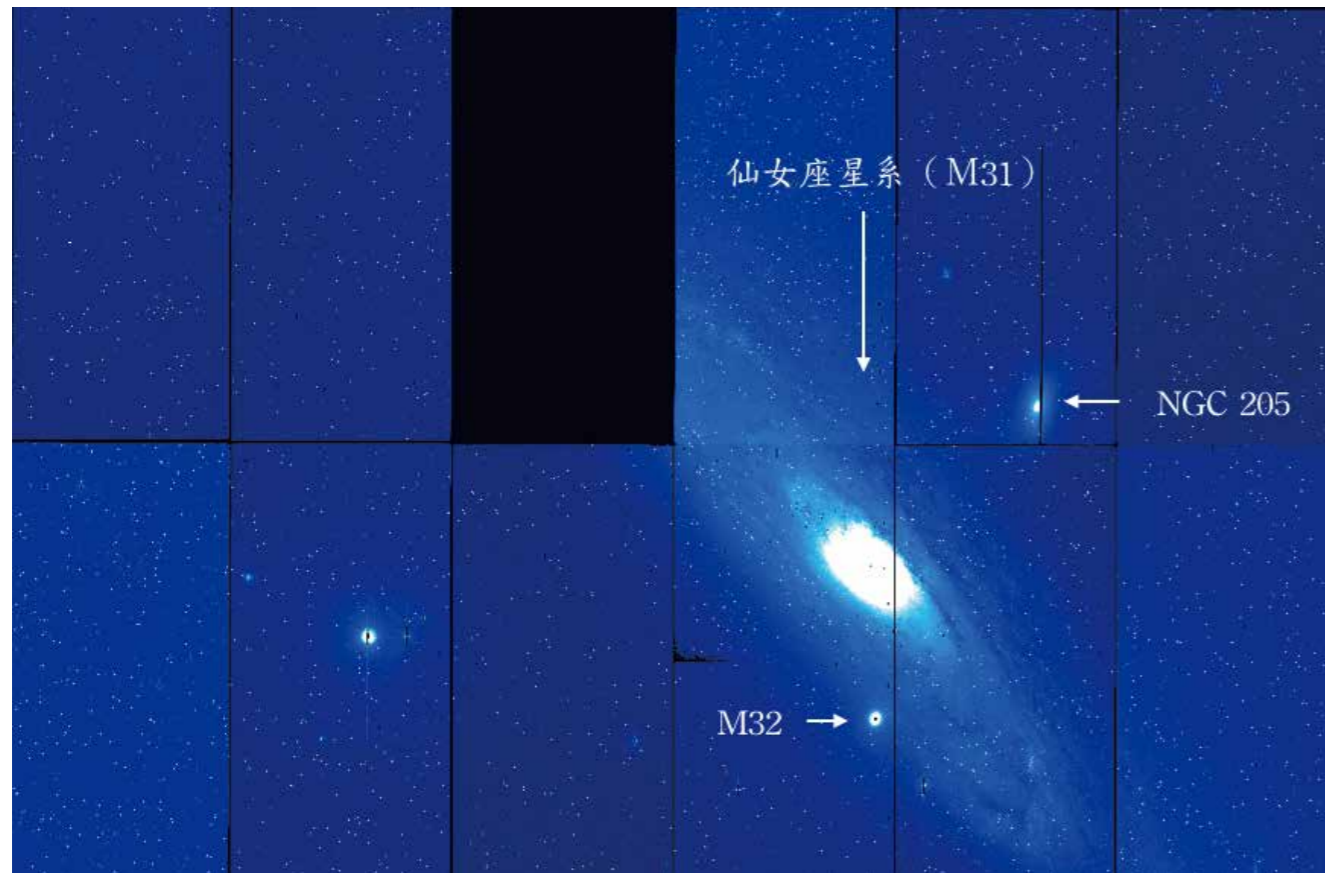
圖一：造父變星光變曲線特徵示意圖。

個物體高速遠離我們的時候，它所發出的波頻率會變低、波長會變長。一個最好的例子就是當鳴笛的救護車高速遠離駛去時，它的聲波頻率就變低了，同樣的原理在觀測遙遠的星系也適用。因為高速離我們遠去，所以在觀察它們的光譜時，會發現光譜往長波長（也就是偏向紅光）的方向移去，這個現象在天文學中稱之為紅移。透過紅移的程度，我們便可以測定星系的距離。但是光譜觀測需要耗費相當多的望遠鏡時間，因此天文學家們便想到使用光度的方式來測紅移。主要的原理是建

立不同星系在不同紅移值的光度模板，然後將觀測所測得的光度與模板擬合，藉此標定光度紅移。但是過去的研究顯示，不同的模板所測出的光度紅移存在相當程度的系統誤差。為了減少誤差，天文學家們想到使用機器學習法來量測光度紅移。

## 使用機器學習進行全自動觀測與後續追蹤

但是在單純地對天體進行分類之餘，機器學習在巡天計畫裡，尤其是後續觀測上，更扮演了重要的關鍵角



圖二：帕洛馬瞬變天體巡天計畫相機視野極廣，不僅仙女座星系，連附近的矮星系 M32 以及 NGC 205 都能一併入鏡，綽綽有餘！（NASA/JPL/IPAC）

色。最有名的例子便是由加州理工學院主導，許多國際機構（包括臺灣的中央大學與清華大學）所參與的帕洛馬瞬變天體巡天計畫（Palomar Transient Factory, PTF）。這個計畫的主力是帕洛馬天文台裡兩座老舊且乏人問津的一米等級施密特望遠鏡。施密特望遠鏡視野廣，曾是上世紀初觀測天文學的主流。但折射式的設計相當佔空間，加上帕洛馬天文台還有更新的五米望遠鏡，這兩座老舊天文台瀕臨汰舊的命運。

既然無人使用，何不讓它們專心進行巡天計畫呢？加州理工學院於是從加法夏天文台購入汰換掉的舊相機，安裝在口徑一米二的 P48 望遠鏡上，賦予這座望遠鏡新的生命。

雖然加州光害嚴重，最暗頂多只能看到 20 星等的天體，但是對於研究明亮的瞬變天體綽綽有餘。加法夏天文台的舊相機安裝在施密特望遠鏡上，將其廣角視野的特性發揮得淋漓盡致，單幅影像可達到 7.2 平方度，相當於 40 倍滿月的大小，連離地球最近的仙女座星系都能完全入鏡（圖二）。廣角相機雖然可以快速掃瞄大部分的天區，但缺點是解析度低，一個像素約一個平方角秒（一般天文相機為 0.25 ~ 0.01 平方角秒），因此需要高解析度的相機進行後續觀測。這時候帕洛馬山頭上另一座老舊的 1.5 米望遠鏡——P60 便派上用場了，為了讓這座發揮望遠鏡的最大效益，同時取得不同波段的光度資訊，加州理工大學特別設計了一具多波段測光暨光譜儀

（軟體部分由臺灣中央大學主導），可以一次取得 griz 四個可見光濾鏡的光度，同時還能拍攝低解析度的光譜，對於新發現的瞬變天體進行快速且準確地分類。

但是問題來了，巡天相機一個晚上偵測到的瞬變天體數量極多，如果一個一個目視分類，即便有一組天文學家日以繼夜投入也分類不完，更遑論瞬變天體稍縱即逝，如何能在極短的時間內挑出有趣的天體，並啟動後續觀測呢？聰明的天文學家們想到一個方法，那就是利用機器學習自動分類，並依重要性排序。這麼一來，便能達到幾乎即時（在同一個晚上）啟動後續觀測，重要的是天文學家們還不用熬夜！

#### 公眾科學家協助處理巨量資料

不過機器學習也並非沒有缺點，最大的挑戰是要有一組精心挑選的訓練資料，讓機器利用演算法學習分辨天體的重要性。再者，因每個巡天計畫的軟硬體設計都不一樣，所以針對不同的巡天計畫，需要個別量身定做一組訓練資料。這對於未來的巡天計畫是個致命的缺點，尤其天文學家們累積了足夠的經驗，不斷更新硬體。未來的大型計畫巡天速度更快，資料量更加快速得成長，天文學家們根本無法在短時間內準備好訓練資料。比如正在建造中的大型綜合巡天望遠鏡計畫（Large Synoptic Survey Telescope, LSST），當資料量大到天文學家們無法消化時，又該怎麼辦呢？

其實仰望夜空從來不是天文學家的專利，業餘天文學家一直是天文研究不可或缺的一部分。以往的業餘天文學家多是藉由使用較小口徑的望遠鏡，在自家後院觀測（又稱後院天文學家），協助觀測離我們較近也較亮的天體，例如超新星、新星、小行星以及微重力透鏡等。但是近年來因為大型巡天計畫不斷公開巨量的資料，現在普羅大眾即便手邊沒有望遠鏡，還是能透過網路參與天文計畫，最有名的一個例子是星系動物園（Galaxy Zoo）。

星系動物園開始於 2007 年，最早因史隆巡天的資料量太大，想要依賴公眾科學家之力，來目視分類上百萬個星系。一開始的分類很簡單，主要是將所有的星系影像簡單的分類為橢圓星系、螺旋星系以及合併星系。同時天文學家們也想藉由這個機會，了解公眾科學家與專業天文學家們對天體進行的目視分類是否有所差異。沒想到經過簡單的訓練與篩選，公眾科學家分類的結果與專業天文學家相當一致。更重要的是公眾科學家踴躍參與，在計畫公布的第一天，平均每小時收到 7 萬筆辨識結果。而計畫執行一年後，總共有 15 萬名公眾科學家參與，累計收到 5000 萬筆辨識結果。這讓天文學家們對公眾科學計畫的信心大增。如今星系動物園已來到第四版，除了史隆巡天計畫外，還加入了哈柏太空望遠鏡的影像、英國紅外相機深度巡天計畫以及暗能量相機巡天計畫的影像。而分類的任務，也從一開始的簡單分類，到計算螺旋星系旋臂數量，以及星系核心的大小等更細膩的特徵。而公眾科學計畫所分類的巨量資料，也提供天文學家測試機器學習所不可或缺的訓練資料。

#### 結語

有許多人害怕機器學習，尤其許多天文台的工作人員，認為這會奪走大家的工作機會。其實恰恰相反，機器學習幫我們免去無聊的繁瑣工作，讓天文學家們能集中心力，專注在更重要也更有意思的科學研究上。

此篇文章為 Artificial intelligence in research, *Science*, Volume 357, Issue 6346, page 28-30, 2017. 一文的衍伸

#### 延伸閱讀

1. J. W. Richards *et al.*, Construction of a Calibrated Probabilistic Classification Catalog: Application to 50k Variable Sources in the All-Sky Automated Survey, *The Astrophysical Journal Supplement*, Vol. 203: 27, 2012.
2. S. Dieleman, K. W. Willett and J. Dambre, Rotation-invariant convolutional neural networks for galaxy morphology prediction, *MNRAS*, Vol. 450: 1411-1459, 2015.
3. M. C. Kind, and R. J. Brunner, Exhausting the Information: Novel Bayesian Combination of Photometric Redshift PDFs, *MNRAS*, Vol. 442, 3380, 2014.