預測、學習、與賽局

線上學習簡介

作者：李彥寰

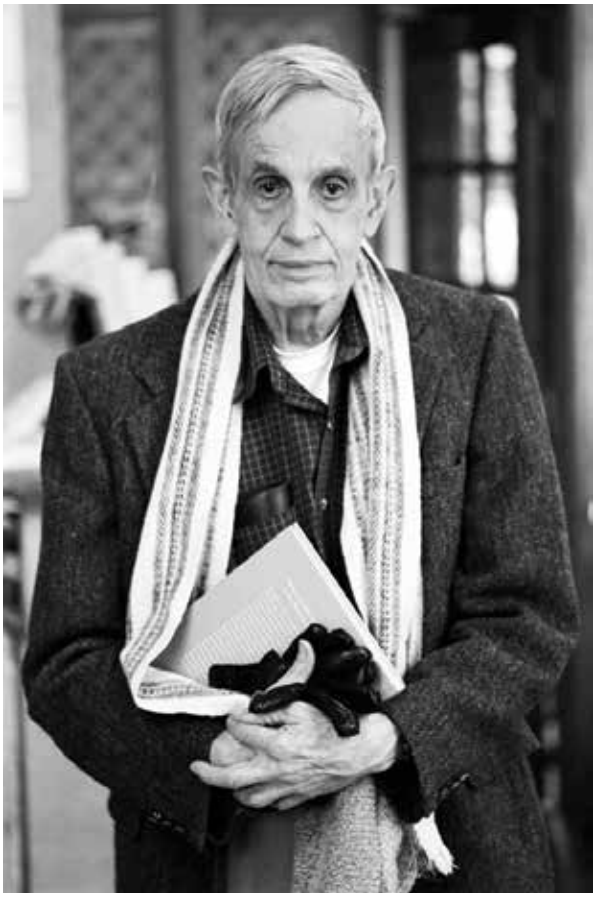
作者簡介：李彥寰是國立臺灣大學資訊工程學系暨資訊網路與多媒體研究所及數學系合聘助理教授，研究領域是機器學習和最佳化演算法的設計與分析。

（Gerd Altman，Pixabay）

線上學習（online learning）是當代機器學習理論的主流研究方向之一。最近幾年，在機器學習最具代表性的純理論會議 COLT（Conference on Learning Theory），總能看到許多的線上學習論文，有興趣的讀者可以在 YouTube 上找到這個會議的官方演講錄影。線上學習在實作上也非常重要。線上學習演算法的每一次迭代只需要讀取一筆資料，因此特別適用於資料量大到難以一次全部讀取的情境 —— 這就是當代大部分機器學習問題的典型情境。線上學習和賽局論之間有著非常密切的關係；例如在一些賽局之中，如果所有玩家都採取線上學習的策略，那麼他們出的招法將收斂至納許均衡（Nash equilibrium）。

回顧統計學習理論

什麼是線上學習？討論這個問題之前，讓我們先快速回顧機器學習的標準理論：統計學習理論（statistical learning theory）。以圖像分類（image classification）爲例。假設我們的訓練資料（training



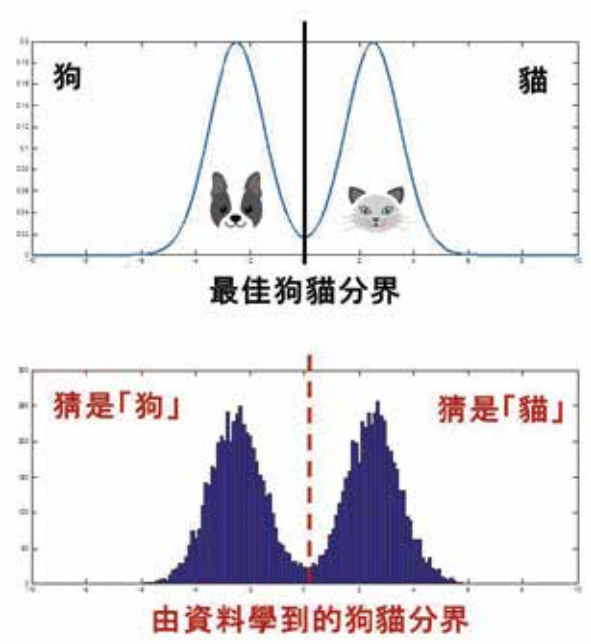
納許（John Nash，1928~2015）。（維基，Peter Badge）

data）是一大堆貓狗照片及對應的標籤，標籤標示照片裡的動物是狗還是貓。圖像分類的目標是設計一個機器學習演算法，這個演算法的輸入是訓練資料，輸出則是一個分類器（classifier）。我們希望這個分類器遇到任何沒看過的測試資料（test data）時，能夠只憑照片就準確預測其標籤。

在理論上刻畫這個圖像分類問題並不簡單。首先，訓練資料和測試資料必須「看起來不一樣，但本質上一樣」。如果所有的照片和標籤都一模一樣，那麼我們輸入演算法的其實只有一筆資料，演算法看過的例子太少，分類器的表現大概很糟糕。如果訓練資料和測試資料非常不同，那麼演算法看過的訓練資料並無法幫助它認識測試資料，分類器的表現大概還是很糟糕。另外，我們合理預期訓練資料愈多，分類器愈準確。好的理論應該要能說明以上這些現象。

利用機率論，統計學習理論提供了一個很巧妙的機器學習的理論架構。在統計學習理論中，我們做以下假設：

1. 訓練資料和測試資料都是隨機的，所以「看起來不一樣」。
2. 訓練資料和測試資料都服從同樣的機率分布，所以「本質上一樣」。
3. 訓練資料和測試資料是統計獨立的，所以「資料量愈多，分類器愈準確」。



圖一。

前兩個性質應該很明顯，讓我們著重討論第三個性質。見圖一。爲了圖示方便起見，我們假設每張狗貓照片可以表示爲數線上的一個點。如果訓練資料是統計獨立的，根據大數法則，當資料量很大時，資料的直方圖（histogram）應該要非常貼近眞正的機率分布，因此我們可以由訓練資料推估出一

個不錯的「狗貓分界點」，建立一個分類器；如果資料量不夠大，直方圖可能和眞實的機率分布相距甚遠，那麼根據直方圖建立的分類器準確率就低。使用比較進階的數學理論，我們甚至可以推導出錯誤率隨著資料量下降至零的速度。因爲測試資料也是隨機的，這裡錯誤率的定義是猜錯其標籤的機率。

統計學習理論在過去幾十年間取得巨大的成功。知名的機器學習演算法 —— 支持向量機（support vector machine）—— 就是基於統計學習理論設計的，目前針對深度學習的理論研究也都是基於統計學習理論。但是，如果拿掉「統計獨立性」的假設，這套理論就不適用了。考慮一個極端的例子：如果訓練資料裡的每一筆都和第一筆一模一樣，那麼不管資料量多麼大，直方圖都無法反映眞實的機率分布，那麼分類器的準確率必然極低。

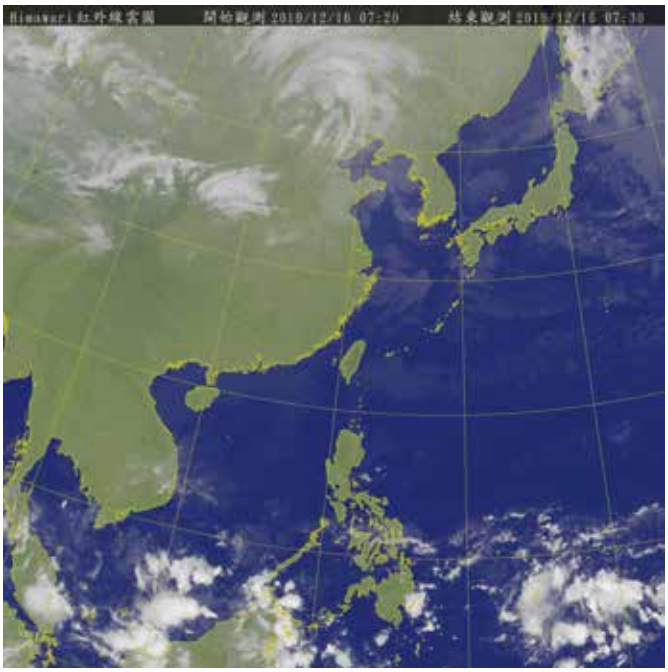
在圖像辨識這類問題中，「統計獨立性」看起來是個可以接受的假設，但在其它問題中不一定如此。例如天氣預報問題：我們想要設計一個演算法，它可以根據過去每日晴雨的紀錄，輸出爲明日是晴是雨的預測。根據常識，過去每日的晴雨可以視爲隨機，但它們似乎並不統計獨立；梅雨季往往一連幾天都是綿綿細雨；颱風季通常要不大晴天，要不大風大雨。那麼大數法則不適用，我們該怎麼辦呢？線上學習是一招！

線上學習的基本概念

線上學習的想法非常激進：讓我們拋棄機率論吧！在線上學習的數學模型中，我們不假設訓練資料和測試資料爲隨機產生，但它們具有「不確定性」；資料可能是隨機產生的，也可能是個懷有惡意的敵手（adversary）給出的。這個「不確定性」比隨機性還更難以捉摸，畢竟資料背後並沒有一個機率分布，我們無法說出期望值、變異數等等量化指標。芝加哥學派的著名經濟學家奈特（Frank Knight）在其名著《風險、不確定性與利潤》（*Risk, Uncertainty, and Profit*）中，就說過：「可以（用機率和期望值）量化的風險和不確定性很不一樣，相較之下前者根本不是一種不確定性。」有興趣的讀者可以自行在網路上搜索「Knightian uncertainty」這個關鍵字。

在統計學習理論中，我們可以用對於測試資料的猜錯機率、損失的期望值等等來衡量一個機器學習演算法的好壞，這看來十分自然。在線上學習理論中，沒有機率和期望值，我們如何衡量一個演算法的好壞呢？此處有一個重要的想法轉折：既然我們無法用機率預測未來，那麼我們就回首過往吧！如果每次回首過往，都覺得一個機器學習演算法的表現愈來愈好，那麼我們就說這個機器學習演算法有「學到東西」。

要回首過往，必然要在機器學習的過程中引入一個時間的概念。線上學習將學習的過程視爲一個序列賽局（sequential game）。以天氣預報爲例。在這個天氣預報的序列賽局中有兩個玩家：一個叫做預報員，另一個叫做環境。這個序列賽局有多個回合。每個回合的開始對應到現實世界中的晚間新聞時段，結束時間爲隔天的晚間新聞之前。在每回合的開始，預報員宣布對於明天會不會下雨的預測；到了現實生活中的「明天」，環境宣布當天的天氣，預報員因此知道昨日的猜測是對或錯，然後結束一



（交通部中央氣象局）

回合。對於預報員而言，每天都要根據訓練資料——過去每日晴雨的紀錄——對明日的天氣作出預測，眞實發生的明日天氣就是測試資料；每過完一回合，昨日的測試資料就變成明日的訓練資料。如果每次預報員回首過往，總是發現自己的預報比上次回首過往時看起來更準，就可以宣稱自己從資料中學到一些東西。

順帶一提，因爲考慮的總是一邊收進新資料一邊做預測的問題，所以這門學問被稱爲線上學習。

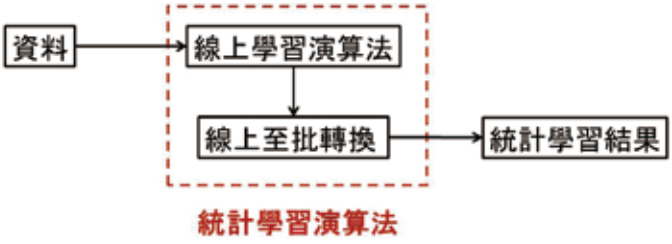
後悔值

怎麼樣叫做「每次回首過往，都比以前更準」呢？這個問題並不簡單。在天氣預報問題之中，預報員必須在環境宣布天氣前給出氣象預報，因此環境可以根據播報員的氣象預報決定明日是什麼天氣。環境可以用一個簡單的策略，讓預報員每次都預報錯誤：當預報員說隔天是晴天時，就讓隔天下雨；當預報員說隔天是雨天時，就讓隔天放晴。在這個時候，預報員不管採用什麼演算法，每次回首過往，看到的永遠是預報全部失敗的紀錄，怎麼可能會覺得自己比以前更準呢？

因此，線上學習關注的並不是達到永遠正確的預測，而是達到很小的後悔值（regret）。以上一段的例子來說，既然不管使用什麼機器學習演算法都將預測失敗，我們並不會對「始終預測失敗」這個結果感到後悔，因此後悔值爲零。準確來說，假設過完了 t 天，後悔值的定義爲：機器學習演算法在這 t 天內猜錯天氣晴雨的次數，減掉在這 t 天內「最佳預測策略」猜錯天氣晴雨的次數。如果隨著天數增加，一個線上學習演算法的平均後悔值（後悔值除以天數）趨近於零，那麼這個演算法就算是隨著時間不斷在進步；此時我們稱這個演算法「無悔」（no-regret）。線上學習這個領域的研究，簡單來說就是爲各種線上學習問題設計無悔的演算法。

「最佳預測策略」這個詞值得推敲。首先，其定義視你考慮的所有候選策略的集合而定。如果你考慮的候選策略的集合愈大，設計無悔演算法這個任務就變得愈困難。大部分研究考慮的是所有最佳的「不變」策略。以天氣預報問題爲例，所有的不變策略只有兩個：永遠預測晴天以及永遠預測雨天。另外，最佳預測策略會隨著時間變化。再以天氣預報問題爲例，如果前十天都下雨，那麼對於前十天，最佳的不變策略就是永遠預測下雨；如果之後

每天都放晴，那麼到了第二十一天，最佳的不變策略就成了永遠預測晴天。最後，我們只有在回首過往時才知道最佳策略爲何。因此，如果一個演算法無悔，某種程度上我們可以說這個演算法具有「未卜先知」的能力。



圖二。

線上學習演算法的統計學習性質

既然線上學習的理論可以處理任何資料，假設資料滿足統計學習理論的假設（統計獨立且服從相同的機率分布），那麼一個無悔的線上學習演算法可能會有好的統計學習性質。確實如此！這在文獻中稱爲線上至批轉換（online-to-batch conversion）。在滿足一些理論假設的前提下，線上至批轉換的內容爲：將一個線上學習演算法每回合的輸出做個平均，這個平均輸出的統計誤差將不大於後悔值除以回合數。因此，如果一個線上學習演算法無悔，其平均輸出的統計誤差將隨著回合數增加趨近於零。

藉著線上至批轉換，我們可以用線上學習的想法設計統計學習演算法。這樣造出來的統計學習演算法往往有個特點：每一次迭代只需要一筆資料，計算複雜度與資料量無關，特別適合用於資料量特別大的機器學習問題。資料量特別大正好是當代許多機器學習問題常見的狀況。例如在深度學習領域著名的演算法 AdaGrad，原本作爲線上學習演算法被提出，經過線上至批轉換才變成一個統計學習演算法，在深度學習領域被廣泛使用。這已經是機器學習領域很常用的招數了。以後你看到一個「統計學習演算法」時，不妨想想那是否如圖二所示，是個披著統計學習外表的線上學習演算法。

對於天氣預報問題，細心的讀者可能會有個疑問：爲什麼我們容許環境「慢出」？對於後悔值定義中的「最佳策略」，或許也有讀者想問：爲什麼大部分的研究只考慮最佳的不變策略？這兩個問題可以從線上至批轉換的角度來回答。如果我們不容許環境慢出，線上至批轉換無法成立；如果我們不只考慮最佳的不變策略，線上至批轉換無法定義。更詳細的解說需要用到機率論中鞅（martingale）的概念。有興趣的讀者可以看看切薩畢安奇（Nicolò Cesa-Bianchi）教授等人寫的〈線上學習演算法的推廣能力〉（On the generalization ability of on-line learning algorithms）及相關文獻。

MWU 演算法

愛麗絲想從賭馬賺錢。她對賭馬不熟悉，不知道哪些馬值得下注。幸好她有一群朋友，她可以觀察在每次比賽中每個人的押注狀況，並藉由這些觀察決定她自己的賭馬策略。這些朋友中有些人是賭馬專家，但是她目前不知道這些人是誰。請幫愛麗絲設計一個賭馬演算法，讓愛麗絲賺的錢和（未知的）賭馬專家差不多。

以上是弗羅因德（Yoav Freund）教授和夏皮爾（Robert Schapire）博士在他們的經典論文〈線上學習的決策理論推廣與提升學習的應用〉（A decision-theoretic generalization of on-line learning and an application to boosting）所提出的問題。具體來說，讓我們考慮一個多回合的雙人賽局，玩家爲愛麗絲和環境，一個回合對應至一場賽馬賭局。每回合一開始，愛麗絲根據隨機選一位朋友，照著這位朋友的方式下注，然後環境宣布每位朋友輸多少錢（贏錢視爲輸負數的錢）。愛麗絲在每一回合的損失爲她輸錢的期望值。目標是設計一個無悔的演算法，以決定每次選擇朋友時用的機率分布。在這裡，後悔值的定義是愛麗絲的總損失減去輸錢最少的朋友的總輸錢數；如果後悔值除以回合數趨近於零，我們就說這個演算法無悔。

賭馬問題的後悔值定義和天氣預報問題中的定義不大一樣，但概念是相同的：「輸錢最少的朋友」在前十回合和前二十回合可能是不同人；除非愛麗絲未卜先知，不然她不會知道誰將是未來輸錢最少的朋友；如果愛麗絲使用一個無悔的演算法，那麼她和輸錢最少的朋友在每回合輸錢的平均差距趨近於零。

弗羅因德教授和夏皮爾博士提出的演算法非常簡單，後來被稱爲乘性權重更新（multiplicative weight update，MWU）演算法。一開始，愛麗絲不知道該選哪位朋友，所以選每位朋友的機率都一樣。在之後的每一回合，愛麗絲選任何一位朋友的機率正比於該朋友在過去的總輸錢數乘以某個適當的正實數（在文獻中稱爲學習率）的指數；這也就是演算法名稱的由來。我們可以看到，如果一位朋友一直輸錢，愛麗絲選他的機率很快地會降到幾乎是零，可以預期最後愛麗絲選到的通常是表現很不錯的朋友。這個演算法無悔的數學證明並不複雜，有興趣的讀者可以找弗羅因德教授和夏皮爾博士的論文來讀。

賭馬並不是非常實際的問題，但這個問題的數學結構出現在很多地方。當你有幾個可能的選擇，不知道該選哪一個的時候，都可以考慮使用 MWU。例如弗羅因德教授和夏皮爾博士提出 MWU 是爲了

1890 年所發行的卡洛爾（Lewis Carroll）所著《愛麗絲夢遊記》童書版書中坦尼爾（John Tenniel）的插畫。（維基）

設計有名的 AdaBoost 演算法，其中 MWU 用來決定每一筆訓練資料的權重。另外，我們可以用 MWU 推導出凸優化（convex optimization）中有名的熵鏡下降（entropic mirror descent）演算法；MWU 的變形還可以推導出金融工程的泛投資組合選擇（universal portfolio selection）演算法和深度學習中被廣泛使用的隨機梯度下降（stochastic gradient descent，SGD）演算法。MWU 可以說是線上學習領域中最經典的演算法之一。

線上學習與賽局

賽局論中，最有名的概念大概是納許均衡了！但是納許均衡的概念並不很直觀；在一些如囚犯困境的例子中，納許均衡甚至看起來不大符合直覺。有沒有什麼方法可以讓納許均衡「自然而然」地出現呢？這是賽局論中「賽局學習」（learning in games）這個子領域主要關注的問題之一。從機器學習的角度來看，這個領域和線上學習驚人地相似，十分有趣。

剪刀、石頭、布。（photoAC）

讓我們考慮這個遊戲。假設輸家要付給贏家一塊錢，可以證明納許均衡爲雙方都用 $(1/3, 1/3, 1/3)$ 的機率出剪刀、石頭、布，此時雙方賺到的錢的期望值都是零。標準的賽局論說法是：「理性的玩家」會照著納許均衡出拳，雙方都預期賺不到錢。

現在考慮多回合的剪刀石頭布。大部分的正常人似乎就不會像上述「理性的玩家」那樣隨機出拳了。對於正常人來說，一個合理的策略是觀察對手過去出拳的模式，預測對手接下來出什麼拳，然後決定自己的出拳。例如觀察到對手每次出布之後一定會出剪刀，那麼看到對手上回合出布，自己這回合一定出石頭。這樣根據歷史資料做預測，和天氣預報本質上是一樣的，也可以看成是個線上學習問題。如果兩個玩家都採用無悔的線上學習演算法，最後結果如何呢？決定自己的出拳。例如觀察到對手每次出布之後一定會出剪刀，那麼看到對手上回合出布，自己這回合一定出石頭。這樣根據歷史資料做預測，和天氣預報本質上是一樣的，也可以看成是個線上學習問題。如果兩個玩家都採用無悔的線上學習演算法，最後結果如何呢？

讓我們說得更具體一些。假設對局雙方爲愛麗絲和巴布。對於愛麗絲來說，她面臨的是這樣一個線上學習問題：每回合一開始，她隨機選擇剪刀、石頭、布三個行動中的其中一個，然後她觀察到巴布出的拳。如果愛麗絲贏了，這回合的損失爲負一元，否則損失爲一元。後悔值的定義則是愛麗絲總

損失的期望值減去用最佳不變的選擇剪刀、石頭、布的機率分布所造成的損失期望值。對於巴布來說，他面臨的線上學習問題的數學架構一模一樣，只是對局雙方對換。

對於像剪刀石頭布這樣的簡單雙人零和賽局，可以證明，如果愛麗絲和巴布都採用無悔的線上學習演算法，則這兩個人分別用來決定出拳的機率分布的平均值都會收斂到 $(1/3, 1/3, 1/3)$，也就是納許均衡所預測的結果。因此，我們造出了一個相對自然的情境，在這個情境之中納許均衡自然而然地出現了。

可以想見賽局學習這個領域的學者們必然需要研究線上學習演算法。例如愛麗絲和巴布所面臨的問題，都是有幾個可能的選擇（剪刀、石頭、或布），不知道該選哪一個，和 MWU 解決的問題類似，因此也可以用 MWU 類型的演算法來解決。從線上學習的角度來看，其實賽局學習領域的專家們獨立發現了許多重要概念。例如 MWU 在賽局學習領域有其對應的版本，名字叫做謹慎虛擬對局（cautious fictitious play）。

結語

相較於標準的統計學習理論，線上學習是一門相對年輕的、還正在蓬勃發展的理論。線上學習演算法可以處理資料不具備統計獨立性的情況，也可以經由線上至批轉換，變成一個統計學習演算法；這時我們可以用相對低計算複雜度的演算法處理大量資料。

許多線上學習演算法都可以說是 MWU 的特例。有興趣的讀者可以參考阿羅拉（Sanjeev Arora）教授等人所寫的〈乘性權重更新演算法 —— 元算法與應用〉（The multiplicative weights update method: A meta-algorithm and applications）以及凡德霍文（Dirk van der Hoeven）先生等人所寫的〈線上學習中指數權重的多面性〉（The many faces of exponential weights in online learning）這兩篇文章。

線上學習這門學問相對年輕，所以相關書籍不多。這個領域的「聖經」應該算是切薩畢安奇和盧戈西（Gábor Lugosi）這兩位教授所寫的《預測、學習、與賽局》（*Prediction, Learning, and Games*）；這本書寫得非常好，涵蓋的主題和參考文獻都很詳盡，可惜出版時間在 2006 年，缺少近十幾年來的許多重要研究成果。如果讀者想知道線上學習這個領域的最新進展，可以看看 COLT、NeurIPS、ICML、AISTATS 這幾個頂級機器學習會議的論文集。

延伸閱讀

▶ Nicolò Cesa-Bianchi 與 Gábor Lugosi，《預測、學習、與賽局》，Cambridge University Press，2006。

▶ D. Fudenberg 與 D. K. Levine，"Learning and equilibrium"，*Annu. Rev. Econ*, Vol. 1:385-420，2009。

▶ http://mindreaderpro.appspot.com/
是根據線上學習演算法開發的小遊戲。