

[首頁](#) / [雜誌](#) / [機器學習贏家祕訣](#)

技術

# 機器學習贏家祕訣

How to Win with Machine Learning

阿傑·艾格拉瓦 Ajay Agrawal, 約書亞·格恩斯 Joshua Gans, 阿維·高德法布 Avi Goldfarb

2020年10月號(2020台灣CEO100強) | 2020/9/24

瀏覽人數：1858

本文出自



2020台灣CEO100強

2020年10月號

看更多雜誌文章



隨著愈來愈多企業運用人工智慧來輔助生產或提供服務，如何靈活運用機器學習，打造長期的競爭優勢？  
進入障礙，與已經搶占市場的先行者分庭抗禮？

**過**去十年間，人工智慧 ( AI ) 有一個令人振奮的領域突飛猛進，那就是機器學習 ( machine learning ) 轉化為各項預測，使亞馬遜 ( Amazon )、蘋果、臉書、Google等科技巨擘有能力大幅改善產品平台，有時甚至與大型科技公司競爭。

以總部位於多倫多的BenchSci為例。這家公司設法加快藥物開發的流程，目標是讓科學家更容易在大海撿內部的資料庫及大量已發表的科學研究中，找到最關鍵的資訊。為了讓候選的新藥進入臨床試驗，科學家BenchSci發現，科學家若是能運用已完成的大量實驗中產生的更好見解，就可以少做一些這類實驗，並獲

BenchSci確實發現，機器學習系統可以閱讀、分類科學研究，然後呈現從這些研究中產生的見解；科學家把藥物進入臨床試驗通常所需的實驗數量減半。更具體地說，他們可以使用這項技術，找到恰當的生物試的必要物質。藉由仔細檢視已發表的文獻來找出這些物質，而不是從頭再發現它們，就可以大幅縮短生產可以省下超過170億美元，在這個研發報酬率已變得極低的產業裡，如此龐大的成本節約可能轉變整個市場；挽救許多生命。

值得注意的是，目前BenchSci在它專精的領域裡所做的事情，其實很像Google在整個網際網路做的事情：使用機器學習在搜尋產業取得領先。Google可以協助你了解如何修理洗碗機，讓你避免大老遠跑一趟圖書館查資料，或者省下昂貴的修理服務費用，同樣的，BenchSci協助科學家找出合適的試劑，而免除了進行大量研究與實驗的麻煩或費用。以前，科學家常用Google或PubMed來搜尋文獻（這過程要花好

幾天)，然後閱讀文獻（也要花好幾天），之後訂購三到六種試劑並進行測試，接下來從中挑選出一種試劑（需要花數週的時間）。現在，他們在BenchSci搜尋幾分鐘，然後訂購並測試一到三種試劑，接著從中挑選一種試劑（進行較少的測試，花費較少個星期）。

許多公司已經開始使用人工智慧來運作，而且知道可以採取哪些實際的做法，把人工智慧整合到本身營運中，善用它的力量。但隨著公司愈來愈熟練這個領域，他們**必須考慮一個更廣泛的議題：如何利用機器學習，在企業的周圍打造一道可防禦的護城河，也就是創造出競爭對手無法輕易模仿的東西？**以BenchSci為例，它最初的成功是否會吸引Google來競爭？如果會的話，它如何維持領先地位？

接下來我們會說明，公司若是以人工智慧賦能的產品或服務進入某個產業，要如何建立可持久的競爭優勢，以及提高進入障礙以阻擋後來者。我們注意到，提早行動通常是一大優勢，但重點不僅止於此。我們會提到，新技術的後期採用者仍然可以藉由找到利基市場而迎頭趕上，或至少縮小一些落差。

## 以AI進行預測

企業使用機器學習來辨識各種形態，接著做出預測，包括：什麼能吸引顧客、什麼能改善營運，或者什麼能協助改善產品。然而，在根據這些預測來制定策略之前，你必須先了解這個預測流程所需的輸入資料、取得那些輸入資料會面臨的挑戰，以及回饋資料如何逐漸改進演算法，以做出更好的預測。

### 輸入並不簡單

在機器學習的情境中，預測是指資訊輸出，這是在輸入資料以執行演算法之後輸出的資訊。例如，你的行動導航應用程式提供有關兩點之間最佳路線的預測，它所用的輸入資料包括交通狀況、速度限制、道路大小和其他因素。接著，它使用一套演算法來預測最快的路線，以及所需的時間。

任何預測流程的關鍵挑戰，都是你必須自己創造訓練資料（例如聘請專家來進行分類），或是從現有來源採購訓練資料（例如健康紀錄）；訓練資料是指，為了獲得合理結果而需要的輸入資料。有些類型的資料很容易從公共來源取得（例如天氣、地圖資訊）。如果消費者認為提供自己的個人資料可得到好處，他們可能也願意提供。例如，Fitbit與蘋果手表（Apple Watch）的使用者讓這兩家公司透過他們穿戴的裝置，收集他們的運動量、卡路里攝取量等指標，以管理他們的健康與健身狀況。

不過，如果取得訓練資料需要許多人合作，而這些人無法因此直接受惠，公司可能就很難取得訓練資料以進行預測。例如，導航應用程式可以藉由追蹤使用者，以及取得使用者提報的資訊，來收集交通狀況的資料。這些資訊讓應用程式找出可能塞車的地點，並向朝那些地方行駛的其他駕駛人發出警訊。但是，已經塞在車陣中的駕駛人參與這件事，不太能獲得直接的好處，而且他們可能會覺得困擾，因為這個應用程式知道他們在任何時刻的位置（而且還可能記錄他們的行動）。如果身陷車陣中的使用者拒絕分享自己的資料，或是關閉他們的地理定位器，這個應用程式預先提醒塞車問題的能力，可能就會受到影響。

另一個挑戰，可能是需要定期更新訓練資料。這不見得一定會是問題；如果預測的基本背景條件維持不變，就不必定期更新訓練資料。例如，放射學是分析人的生理機能，這在不同人的身上、不同的時間點都是一致的。因此，資料累積到某個點以後，訓練資料庫再加入額外紀錄的邊際價值幾乎是零。但在其他情況下，演算法可能需要經常更新資料，納入可反映基本環境變化的全新資料。例如，對於導航應用程式，新的道路或圓環、重新命名的街道，以及類似的改變，長期下來會導致應用程式的預測變得較不正確，除非更新構成最初訓練資料的那些地圖。

### **當心回饋迴圈擴大錯誤參數**

在許多情況下，演算法可以利用回饋資料，不斷地改進；若是把實際的結果，拿來和當初用來產出預測的輸入資料作比較，就可以產生回饋資料。如果在明確定義的界線內可能會出現很大的變化，這種工具就特別有助益。例如，如果你的手機以你的頭像作為解鎖機制，你必須先訓練手機辨識你。但你的臉可能有很大的變化。你可能戴眼鏡或不戴眼鏡，可能換新髮型、化妝，或者變胖或變瘦。因此，如果手機只依賴最初的訓練資料，它預測「你是你」的結果就不是那麼可靠了。但實際的狀況是，你每次解鎖手機時，手機都會用你提供的所有圖像來更新演算法。

在動態環境中，回饋資料無法輕易分類及取得，因此很難創造這種回饋迴圈。以智慧型手機的臉部辨識應用程式為例，如果每次輸入臉部資料的人都是手機的主人，回饋資料才能產生比較好的預測。如果長得夠像的其他人也能解鎖，並持續使用這支手機，手機預測這名使用者是手機主人的結果會變得不可靠。

而且，把偏見導入機器學習中非常容易，這很危險，尤其若是有多個因素都能發揮作用的情況下，更是危險。假設放款人使用人工智慧賦能的流程，來評估貸款申請人的信用風險，考慮他的收入水準、就業歷史、人口統計特性等等。如果這個演算法的訓練資料歧視某個群體（比如有色人種），回饋迴路就會讓這種偏見持續存在，甚至變本加厲，使有色人種在申請貸款時更有可能遭到拒絕。如果沒有仔細定義的參數，以及可靠、無偏見的來源，回饋資料幾乎不可能安全地納入演算法中。

## 在預測中建立競爭優勢

在機器學習領域中建立可長久維持的事業，在很多方面很類似在任何產業中建立可長久維持的事業。你必須先有一個可銷售的產品，盡早在市場上占據一個可防禦的位置，並且讓別人都更難在你之後進入市場。你能否做到這些，取決於你對以下三個問題的答案：

### 掌握資料愈多愈難被超越

**1：你有足夠的訓練資料嗎？**打從一開始，預測機器就必須產生商業上夠好的預測。「夠好」的定義也許可以考慮以下因素：法規（例如，進行醫療診斷的人工智慧必須符合政府標準）、可用性（聊天機器人的運作必須夠順暢，讓打電話進來的人可以回應機器，而不是等著和電話中心的真人說話），或者競爭（試圖進入網路搜尋市場的公司，需要有一定程度的預測準確度，才能和Google競爭）。因此，**創造或取得足夠的訓練資料，以便做出夠好的預測，這個過程所需的時間與精力，就是一種進入障礙。**

這個障礙可能很高。以放射學為例，在這個領域，做預測的機器必須表現得比高技能的真人優異很多，我們才會放心把人命託付給機器。這表示，第一家為放射學打造一般通用人工智慧（可解讀任何掃描圖像）的公司，一開始面臨的競爭很少，因為成功需要大量資料。但如果市場成長迅速，最初的優勢可能維持不久，因為在快速成長的市場中，取得訓練資料的回報可能很大，足以吸引多家財力雄厚的大公司加入。

**科技巨擘無疑擁有先機，但你只要在使用目的和情境上稍微做出一點差異，就可以為你的產品在市場上打造一塊可防禦的領域。**

當然，這表示輸入訓練資料的要求條件，就跟許多東西一樣，受到規模經濟的影響。高成長的市場會吸引投資，隨著時間推移，這會提高下一個新進業者的進入門檻（也迫使已經進入市場的每一家業者，花更多錢去開發或行銷本身的產品）。因此，你有愈多的資料可用來訓練機器，任何競爭對手面臨的障礙就愈大，而這引導至第二個問題。

### **令人望塵莫及的迴圈速度**

**2：你的回饋迴圈有多快？**預測機器運用傳統上屬於人類的優勢，也就是說，它們會學習。機器如果納入回饋資料，就能從結果中學習，並提高下一次預測的品質。

然而，**這種優勢的大小，取決於獲得回饋所需的時間**。以放射檢查為例，如果需要解剖檢驗，才能評估機器學習演算法是否正確預測癌症，這種回饋將會很緩慢；而即使這家公司初期可能在收集及解讀掃描影像方面領先，但它的學習能力會因為回饋慢而受限，因而很難維持領先地位。相反的，**如果獲得預測後可以迅速產生回饋資料，早期的領先就可以轉化為持久的競爭優勢**，因為競爭對手即使是最大的公司，也在

不久之後就無法達到最低限度的有效率規模。

2009年微軟推出Bing搜尋引擎時，Bing獲得微軟全力支持。微軟對它投資了數十億美元。但十多年後，Bing的市占率仍遠遠落後Google，無論是搜尋量還是搜尋廣告收入，都落後Google許多。Bing發現難以追上Google的一個原因，就是回饋迴圈。在搜尋方面，從預測（為一個查詢提供一個頁面，上面有幾個建議的連結）到回饋（使用者點擊其中一個連結）之間的時間很短，通常只有幾秒鐘。換句話說，這個回饋迴圈又快又強大。

Bing進入市場時，Google已經推出以人工智慧為基礎的搜尋引擎十幾年了，已協助數百萬名使用者，而且每天執行數十億次搜尋。每次使用者查詢時，Google便提出最相關連結的預測；接著，使用者從中選擇最好的連結，讓Google能夠更新它的預測模型。這樣就可以在持續擴大的搜尋領域中，不斷地學習。Google有這麼多的使用者提供如此多的訓練資料，所以辨識新事件與新趨勢的速度比Bing還快。最後，快速的回饋迴圈，再加上其他因素（例如Google持續投資龐大的資料處理設備；使用者改用另一種搜尋引擎的實際成本，或他們認為的成本），導致Bing總是落後。其他試圖與Google和Bing競爭的搜尋引擎，甚至連起步的機會都沒有。

## 精準預判顧客需求

**3：你的預測有多準？**任何產品的成功，最終都取決於你付出以後得到什麼。消費者在面對價格一樣的兩種類似商品時，通常會選他覺得品質比較好的那一個。

我們提過，預測的品質通常很容易評估。在放射學、搜尋、廣告和許多其他領域，**公司設計人工智慧系統時，可以採用一個清楚明確的品質標準：準確性。**就像其他的產業一樣，品質最好的產品，需求比較高。不過，以人工智慧為基礎的產品與其他產品不同，因為對大多數其他產品來說，品質較好的產品，價格就較高；銷售劣質產品業者的生存之道，就是使用較便宜的材料，或較便宜的製造流程，然後以較低的價格出售。這種策略在人工智慧環境中是行不通的。**人工智慧是以軟體為基礎，因此提出劣質預測的成本，和優質預測的成本一樣昂貴，因而讓壓低價格的做法不切實際。**如果較佳的預測與較差的預測訂定相同價格，那就沒有理由購買品質較差的預測。



這是Google在搜尋界的霸主地位堅不可摧的另一個原因。競爭對手的預測，常常看起來和Google的很像。在Google或Bing中輸入「天氣」，結果大致相同，也就是天氣預報會先跳出來。但是，如果你輸入一個較不常見的詞，差異就出現了。例如，你輸入「破壞」，Bing的搜尋結果首頁通常是顯示字典的定義，而Google是提供定義，以及「破壞式創新」的研究論文連結。儘管Bing在某些文字的查詢上，表現得跟Google一樣好，但若是查詢另一些文字，Bing在預測使用者想要的內容方面就較不準確。而在其他搜尋類別中，也沒有或只有很少數類別，是大家普遍認為Bing表現比較好的。

## 迎頭趕上

基本上，在人工智慧的領域，如果回饋迴圈快速，而且績效品質明顯，早期的先行者就可以憑藉規模的基礎，來打造競爭優勢。這對後進者來說，意味著什麼？在這三個問題的背後隱藏著一些線索，顯示後來者可採取兩種方式，在市場中找到立足之處。潛在的競爭者不必在這兩種方法中二選一，可以雙管齊下。

### 特殊資料可成致勝關鍵

**找出其他的資料來源，並固守這些來源。**在預測工具的一些市場中，可能會有一些潛在的訓練資料庫，是既有業者尚未掌握的。再以放射學為例，每年有數萬名醫師各自解讀數萬份的掃描檔，這表示仍有數以億計（甚至數十億）的新資料點可用。

先行者是從數百名放射醫師那裡取得訓練資料。當然，一旦他們的軟體開始實際運作，掃描的數量及資料庫裡的回饋數量，都會大幅增加；但是，之前分析及驗證過的數十億份掃描檔，對後來者是一個機會，如果他們匯集那些掃描檔，進行整體分析，就有迎頭趕上的機會。他們若是真的這麼做，或許可以開發出一種能做出夠好預測的人工智慧，並推出上市，之後他們也可以從回饋資料中受惠。

後來者在訓練人工智慧時，也可以考慮使用病理學或剖檢資料，而不是真人診斷的資料。這種策略可讓他們更快達到品質門檻（因為活檢及剖檢比身體掃描更明確），雖然後續的回饋迴圈可能比較慢。

或者，後進者不必去找未使用過的訓練資料來源，而去找回饋資料的新來源，使學習速度比既有業者更快（BenchSci就是這樣做的成功例子）。新進者因率先採用較快速回饋資料的新來源，所以可以從使用者的行為與選擇中學習，以改善產品。不過，在回饋迴圈已經相當快速、既有企業的營運規模已經很大的市場中，新進業者採用這種方法的機會很有限。而且，明顯快很多的回饋，可能顛覆目前的做法，而這表示新進業者不是真的與既有業者競爭，而是取代它們。

## 尋找不同的服務對象

**預測差異化。**另一種可協助後來者變得有競爭力的方法，就是重新定義什麼因素會讓預測變得「更好」，即使只是對某些顧客來說比較好。例如在放射學領域，如果市場需要不同類型的預測，這種策略就有機會實踐。先行者很可能是用來自一個醫院體系、一種類型的硬體或一個國家的資料，來訓練它的演算法。新進者可使用來自另一個系統或另一個國家的訓練資料（以及之後的回饋資料），為那個使用者區隔量身打造人工智慧，只要那個使用者區隔夠明確獨特即可。比方說，如果美國的城市居民與中國鄉下居民的健康狀況不同，用來診斷其中一個群體的預測機器，在診斷另一個群體時可能較不準確。

根據某一類硬體的資料來做預測，也可能提供市場機會，只要那種商業模式可以降低成本，或者可讓更多顧客使用。如今許多放射學的人工智慧系統，是從最廣泛使用的X光機、掃描機、超音波裝置取得資料，而這些設備是由奇異（GE）、西門子（Siemens）和其他根基穩固的製造商所生產。然而，如果把這些演算法用於其他機器的資料，所產生的預測結果可能較不準確。因此，後進者可以針對其他那些機器專門打造產品，找到一個利基市場；如果那些機器的購買或操作成本較低，或是專為特定顧客的需求而製造的，可能就會對某些醫院很有吸引力。

## 你想預測什麼？

預測機器的潛力很大，科技巨擘無疑擁有先機。但請記住，**預測就像經過精確設計的產品，是針對明確的目的與情境而仔細調整過的**。你只要在使用目的和情境上稍微做出一點差異，就可以為你的產品在市場上打造一塊可防禦的領域。儘管魔鬼藏在你如何收集及使用資料的細節裡，你也可以在那裡面找到解救之道。

不過，**在智慧型機器驅動的產業裡，想要與大型科技公司競爭，真正的關鍵在於一個只有人類才能回答的問題：你想預測什麼？**當然，找到答案並不容易。若要找到答案，就必須深入了解市場動態，並仔細分析某些預測的潛在價值，以及根據這些預測而建立的產品和服務有什麼潛在價值。因此也難怪，BenchSci A2系列融資的主要投資者，並不是加拿大本土的科技投資者，而是Google旗下一家專注於人工智慧的創投公司Gradient Ventures。

( 洪慧芳譯自 “How to Win with Machine Learning,” HBR, September-October 2020 )

---

阿傑·艾格拉瓦 Ajay Agrawal

多倫多大學羅特曼管理學院 ( University of Toronto's Rotman School of Management ) 創業與創新講座教授，也是創意解構實驗室 ( Creative Destruction Lab ) 的創辦人。

---

約書亞·格恩斯 Joshua Gans

羅特曼管理學院技術創新與創業講座教授，也是創意解構實驗室首席經濟學家。

---

阿維·高德法布 Avi Goldfarb

羅特曼管理學院AI與醫療保健講座教授，也是創意解構實驗室首席資料科學家。他們三人合著有《預測機器》（ Prediction Machines: The Simple Economics of Artificial Intelligence，繁體中文版書名為《AI經濟的策略思維》）。

---

本篇文章主題 [技術](#)

[⊕ 追蹤主題](#)

---

[延伸閱讀](#)





