

Ch. 16 in Agrawal, A., Gans, J., and Goldfarb, A. (eds.).

The Economics of Artificial Intelligence: An Agenda, 2019, p. 399-422 .

賴孚權節譯

## “Artificial Intelligence, Economics, and Industrial Organization”

by Hal Varian

### 16.1 Introduction

機器學習 (ML) 和人工智慧 (AI) 已經存在了很多年。然而，在過去的五年中，多層神經網路在圖像識別，語音識別和機器翻譯等不同領域取得了顯著進展。人工智慧是一種通用技術，可能會影響許多產業。在本章中，我將考慮機器學習的可用性如何影響提供 AI 服務的廠商和採用 AI 技術的產業的產業組織。我的目的不是要對這個快速發展的領域進行廣泛的概述，而是要對工作中的某些力量進行簡短的總結，並描述一些可能用於未來研究的領域。

### 16.2 Machine-Learning Overview

想像一下，我們有一組數位圖像以及一組描述這些圖像中所描繪內容的標籤，例如貓，狗，海灘，山脈，汽車或人。我們的目標是使用這些數據來訓練計算機，以學習如何預測一些新的數位圖像集的標籤。(有關演示，請訪問 [cloud.google.com/vision](http://cloud.google.com/vision)，以便在其中上傳照片並檢索適合該照片的標籤列表。)

機器視覺的經典方法涉及創建一組規則，以識別圖像中具有人類可識別特徵 (例如顏色，亮度和邊緣) 的像素，然後使用這些特徵來預測標籤。這種「函數化」方法取得的成功有限。現代方法是使用分層神經網路 (layer neural network) 直接處理原始像素。這不僅在圖像識別方面，而且在語音識別，語言翻譯和其他傳統上困難的機器學習任務上，都取得了巨大的成功。如今，計算機可以在許多任務中勝過人類。

這種稱為深度學習的方法需要 (a) 用於訓練的標記數據，(b) 用於神經網路的演算法以及 (c) 用於運行演算法的專用硬體。學術界和科技廠商免費提供了培訓數據和演算法，並且可以用象徵性的費用使用雲端計算設施中的計算時間。

1. 訓練 (用) 數據。例如 OpenImages，這是一個有標記圖像的 950 萬個數據集，而 Stanford Dog 數據集則是 120 個犬種的 20580 個圖像。

2. 演算法。流行的開源 (open source, 或稱開放原始碼) 軟體庫包括 TensorFlow, Caffe, MXNet 和 Theano。
3. 硬體。可通過雲端計算提供商使用 CPU (中央處理單元), GPU (圖形處理單元) 和 TPU (張量處理單元)。這些函數使用戶可以組織大量數據, 這些數據可用於訓練機器學習模型。

當然, 讓專家來管理數據, 調整演算法並培養整個過程也很重要。這些技能實際上是當前的主要瓶頸, 但是大學正迅速面臨挑戰, 即提供創建和利用機器學習所需的教育和培訓。

除了機器視覺, 深度學習研究社群在語音識別和語言翻譯方面也取得了巨大進步。這些領域也能夠取得進步, 而無需以前的 ML 系統所需的那種函數識別。

其他類型的機器學習在此主題的 Wikipedia 條目中進行了描述。機器學習的一種重要形式是增強學習。這是一種學習, 其中機器可以優化某些任務, 例如在國際象棋或視頻賽局中獲勝。強化學習的一個例子是多臂式吃角子老虎機 (multiarmed bandit), 但還有許多其他工具被使用, 其中一些涉及深度神經網路。

強化學習是一種順序實驗, 因此從根本上講是因果關係: 將特定的棋子從一個區分行動到另一個區分會導致獲勝的可能性增加。這不同於僅使用觀測數據的被動式機器學習演算法。

強化學習也可以在對抗環境中進行。例如, 2017 年 10 月, DeepMind 宣布了一種機器學習系統 Alpha Go 0, 該系統通過與自己對戰圍棋賽局開發了一種高效的策略!

「自學式機器學習」模型是一種有趣的賽局論模型。深度網路是否可以完全自己與其他參與者競爭和 或學習合作? 學習的行為會看起來像我們建立的賽局論模型的均衡嗎? 到目前為止, 這些技術已主要應用於完整的訊息賽局。他們會在訊息不完整或不對稱的賽局中工作嗎?

有一個完整的 AI 子區域, 稱為對抗性 AI (或對抗性 ML), 它結合了 AI 的主題, 賽局論和研究攻擊和防禦 AI 系統的計算機安全性。例如, 假設我們有一個訓練有素的圖像識別系統, 平均表現良好。那最壞的情況呢? 事實證明, 有一些方法可以創建對人類無害的圖像, 這些圖像將始終欺騙 ML 系統。就像「視錯覺 (optical illusions)」可以欺騙人類一樣, 這些「ML 錯覺」也可以欺騙機器。有趣的是, 對人和機器的最佳幻覺有很大不同。有關一些示例, 請參見 Goodfellow et al. (2017) 作為說明性示例, 以及 Kurakin, Goodfellow and Bengio (2016) 作為技術報告。計算機科學研究人員已經認識到與賽局論的聯繫。我認為, 該領域提供了許多有趣的合作機會。(參見例如 Sreevallabh and Liu 2017)。

### 16.2.1 What Can Machine Learning Do?

大眾媒體中的機器學習示例強調了新穎的應用, 例如在象棋, 圍棋和乒乓球比賽中獲勝。但是,

也有許多使用機器學習來解決實際業務問題的實際應用程式。Kaggle 是查看 ML 可以解決哪種問題的好地方。該廠商舉辦機器學習比賽。廠商或其他組織提供一些數據，問題陳述和一些獎金。然後，數據科學家使用數據來解決所提出的問題。優勝者可以將獎金帶回家。網站上有超過 200 場比賽。這是一些最新的。

- 乘客威脅。提高國土安全部威脅識別的準確性: 1, 500, 000 美元。
- 房屋價格。提高 Zillow 的房屋價格預測準確性: 1, 200, 000 美元。
- 訪問 Wikipedia 頁面。預測將來訪問 Wikipedia 頁面的流量: 25, 000 美元。
- 個性化醫學。預測基因變異的效果，以實現個性化醫學: 15, 000 美元。
- 出租車旅行持續時間。預測紐約出租車的總行駛時間: 30, 000 美元。
- 產品搜索的相關性。在 homedepot.com 上預測搜索結果的相關性: 40, 000 美元。
- 聚類問題。您可以確定意圖相同的問題對嗎?: 25, 000 美元。
- 子宮頸癌篩查。哪種癌症治療方法最有效?: 100, 000 美元。
- 單擊預測。您能否預測每個用戶將點擊哪些推薦內容?: 25, 000 美元。
- 庫存需求。最大化銷售並最大程度減少烘焙食品的報酬: 25, 000 美元。

令人高興的是，這些都是真實的問題，而組織希望從真實的問題中獲得真正的答案，這才是真正的金錢。Kaggle 給定了如何將機器學習應用於實際業務問題的具體示例。<sup>1</sup>

## 16.2.2 What Factors Are Scarce?

假設您要在組織中部署機器學習系統。第一個要求是要有一個收集和組織感興趣的數據的數據基礎結構數據管道。例如，零售商將需要一個系統，該系統可以在銷售點收集數據，然後將其上載到計算機，然後計算機可以將數據組織到數據庫中。然後，該數據將與其他數據（例如庫存數據，物流數據以及可能有關客戶的訊息）組合在一起。構建這種數據管道通常是構建數據基礎結構中最耗費人力和最昂貴的部分，因為不同的廠商通常具有難以互連的特有的舊系統。

---

<sup>1</sup>披露：我一直是 Kaggle 的天使投資人，直到 2017 年中期被 Google 收購。從那時起，我對廠商沒有任何財務利益。

整理好數據後，就可以將它們一起收集到數據倉庫中。通過數據倉庫，可以輕鬆使用可以操縱、可視化和分析數據的系統。

傳統上，廠商運行自己的數據倉庫，這不僅需要購買昂貴的計算機，而且還要求人工系統管理員保持所有函數正常運行。如今，在諸如 Amazon Web Services, Google Cloud Platform 或 Microsoft Azure Cloud 之類的雲端計算設施中存儲和分析數據越來越普遍。

雲端提供商負責管理和更新託管數據庫和數據分析工具所需的硬體和軟體。從經濟角度來看，有趣的是，以前對用戶（數據中心）固定的成本現在變成了可變成本（數據中心上的租用時間）。組織可以購買幾乎任何數量的雲端服務，因此，即使是小型廠商也可以從最低級別開始，並根據使用情況付費。雲端計算比擁有自己的數據中心更具成本效益，因為可以根據需要購買計算和數據資源。不用說，當今大多數科技創業廠商都使用雲端提供商來滿足其硬體、軟體和網路需求。

雲端提供商還提供各種機器學習服務，例如語音識別，圖像識別，翻譯等。這些系統已經由賣方培訓，可以由客戶立即使用。每個廠商都不再需要為這些任務開發自己的軟體。

雲端提供商之間的競爭非常激烈。提供高度詳細和特定的圖像識別函數，每張圖像的成本僅為十分之一美分或更少，並且在此價格基礎上提供批量折扣。

用戶也可能具有與其自己的業務相關的特有數據，例如上述銷售點數據。與實施流行的機器學習演算法相比，雲端提供商還提供了最新的，高度優化的硬體和軟體。這使用戶可以立即使用強大的工具... 只要他們具備使用它們的專業知識即可。

如果有硬體，軟體和專業知識可用，則只需標記的數據即可。有多種獲取此類數據的方法。

- 作為營運副產品。想想一連串飯店，其中一些飯店的業績要好於其他飯店，而管理層可能會對與業績相關的因素感興趣。上述 Kaggle 競賽中的許多數據是日常營運的副產品。
- 網頁抓取。這是從網站提取數據的常用方法。關於數據收集及其使用方式到底允許什麼存在法律辯論。辯論太複雜，無法在此處討論，但是 Wikipedia 關於 Web 抓取的條目很好。一種替代方法是使用其他人抓取的數據。例如，Common Crawl 數據庫包含在八年的 Web 爬網過程中編譯的 PB 級數據。
- 提供服務。Google 開展語音識別工作時，沒有專業知識，也沒有數據。它僱用了專家，他們提出了語音輸入電話簿的想法，以作為獲取數據的一種方式。用戶將說「Joe's Pizza, University Avenue, Palo Alto」，系統將以電話號碼回應。數位化的問題和由此產生的用戶選擇被上傳到雲端中，機器學習被用於評估 Google 答案和用戶操作之間的關係，例如，撥打建議的電話號碼。機器學習訓練使用了來自數以百萬計的個人號碼請求的數據並

迅速學習。ReCAPTCHA 應用了類似的模型，其中人類標記圖像以證明它們是人類而不是簡單的機器人。

- 僱用人員來標記數據。可以使用 Mechanical Turk 和其他系統向人們支付標籤數據的費用 (請參閱 Hutson 2017)。
- 從提供商處購買數據。有許多提供各種數據的提供商，例如郵件列表，信用評分等。
- 共享數據。共享數據對雙方可能是互利的。這在學術研究人員中很常見。開放圖像數據集包含由大學和研究實驗室提供的約 900 萬張帶標籤的圖像。出於多種原因 (例如對公共安全的關注)，可能要求共享。例如飛機上的黑匣子或流行病的醫療數據。
- 來自政府的數據。政府，大學，研究實驗室和非政府機構可以提供大量數據。
- 來自雲端提供商的數據。許多雲端提供商還提供公共數據存儲庫。例如，請參閱 Google 公共數據集，Google 專利公共數據集或 AWS 公共數據集 (Amazon Web Services Public Data sets)。
- 計算機生成的數據。前面提到的 Alpha Go O 系統通過與自己進行圍棋賽局來生成自己的數據。Machinevision 演算法可以使用「合成圖像」進行訓練，這些圖像是通過各種方式行動，旋轉和縮放的實際圖像。

### 16.2.3 Important Characteristics of Data

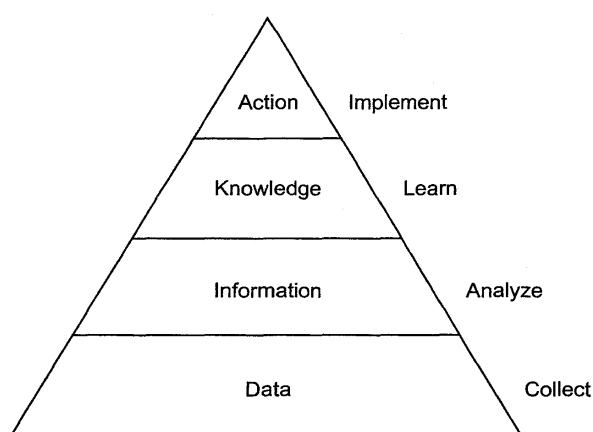


Figure 16 . 1 : 訊息金字塔。

訊息科學使用「數據金字塔」的概念來描述數據，訊息和知識之間的關係。某些系統必須收集原始數據，然後組織和分析該數據以將其轉變為訊息，例如人類可以理解的文本文檔圖像。想

一想圖像中的像素變成了人類可讀的標籤。過去，這是由人類完成的 將來，越來越多的事情將由機器來完成。(見圖 16 . 1。)

這些來自訊息的洞察力可以轉化為知識，通常體現在人類中。我們可以認為數據存儲在位元中，訊息存儲在文檔中，知識存儲在人類中。對於訊息 (書籍，文章，網頁，音樂，視頻) 和知識 (勞動力市場，顧問)，存在發達的市場和管制環境。從無組織的位元集合的意義上講，數據市場並不發達。也許這是因為原始數據通常在很大程度上取決於上下文，並且在轉化為訊息之前不是很有用。

### *Data Ownership and Data Access*

有人說「數據是新的石油」。當然，它們在一個方面是相似的：兩者都需要完善才能有用。但是有一個重要的區別：石油是一種私人物品，石油的消費是競爭者：如果一個人消費石油，那麼其他人的石油消費就更少了。但是數據是非敵對性的：一個人對數據的使用不會減少或減少另一個人的使用。

因此，我們實際上應該考慮數據使用，而不是只關注數據「所有權」(適用於私人物品的概念)。數據很少像私人產品出售那樣被「出售」，而是被許可用於特定用途。目前，歐洲正在就「誰應該擁有自動駕駛汽車數據」進行政策辯論。一個更好的問題是問「誰應該使用自動駕駛數據，他們可以做什麼？」這種表述強調了許多方面可以同時使用自動駕駛車輛數據。實際上，從安全角度考慮，很可能應該允許多方使用自動駕駛汽車數據。汽車中很容易有幾個數據收集點：引擎，導航系統，駕駛口袋裡的手機等等。在沒有充分理由的情況下要求排他性將不必要地限制可以對數據執行的操作。

Ross Anderson 對發生飛機失事時會發生什麼的描述，很重要，說明了為什麼允許多方使用數據可能很重要。

飛機墜毀時，這是頭版新聞。調查人員組成的小組趕到現場，隨後的詢問是由利益攸關各方的組織的專家進行的，這些組織的利益廣泛，包括承運人，保險人，製造商，飛行員聯盟和地方航空當局。他們的發現由新聞記者和政客檢查，在飛行員的討論中討論，並由飛行教員傳承。簡而言之，飛行社群具有強大而制度化的學習機制。

我們是否不應該需要用於自動駕駛汽車的同類學習機制？某些訊息可以受版權保護。但是在美國，電話號碼簿之類的原始數據不受版權保護。(請參閱有關 Feist Publications, Inc 訴 Rural Telephone Service Co. 案的 Wikipedia 條目)

儘管如此，數據提供者可能會編譯一些數據並按一定條款向其他方提供許可。例如，有幾家數據廠商將美國人口普查數據與其他類型的地理數據合併，並提供許可。這些交易可能禁止轉

售或重新許可。即使沒有受保護的知識產權，契約的條款也構成了可以與其他任何私人契約一樣由法院強制執行的私人契約。

*Decreasing Marginal Returns*

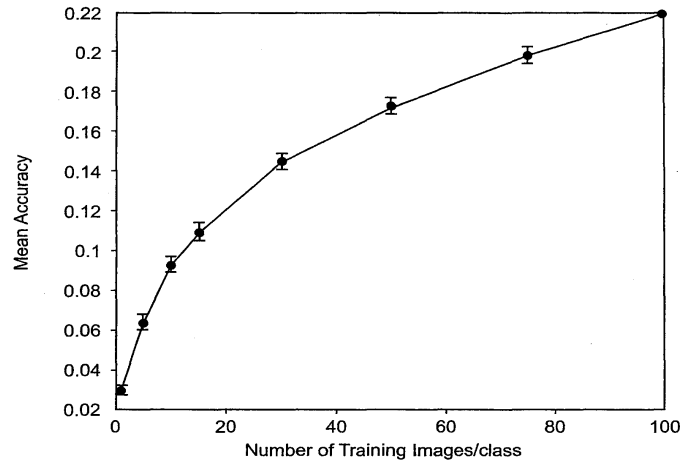


Figure 16 . 2 : 經濟部門採用機器學習。來源:

<http://vision.stanford.edu/aditya86/ImageNetDogs/>

最後，重要的是要理解，數據通常會像其他任何生產要素一樣呈現出規模收益遞減的趨勢。相同的一般原理適用於機器學習。圖 16 . 2 顯示了斯坦福犬品種分類的準確性如何隨訓練數據量的增加而變化。正如人們所期望的那樣，精度隨著訓練圖像數量的增加而提高，但是卻以降低的速度提高。

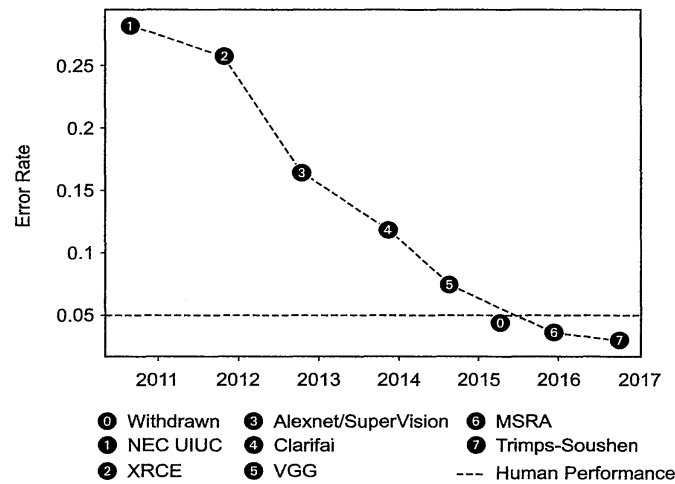


Figure 16 . 3 : Imagenet 圖像識別。來源: Eckersley and Nasser (2017)

圖 16 . 3 顯示了 ImageNet 競賽中的錯誤率在過去幾年中如何下降。關於此競賽的一個重要事實是，在此期間，培訓和測試觀察的數量是固定的。這意味著獲勝系統的性能提高不能一直

取決於樣本量，因為它一直是不變的。其他因素，例如改進的演算法，改進的硬體和改進的專業知識，比訓練數據中觀察的數量要重要得多。

### 16.3 Structure of ML-Using Industries

與任何新技術一樣，機器學習的出現提出了一些經濟問題。

- 哪些廠商和產業將成功採用機器學習？
- 在採用時間和有效使用 ML 的能力上，我們會看到異質性嗎？
- 後期採用者可以模仿早期採用者嗎？
- 專利，版權和商業秘密的作用是什麼？
- 地理在採用模式中的作用是什麼？
- 早期成功採用者是否具有巨大的競爭優勢？

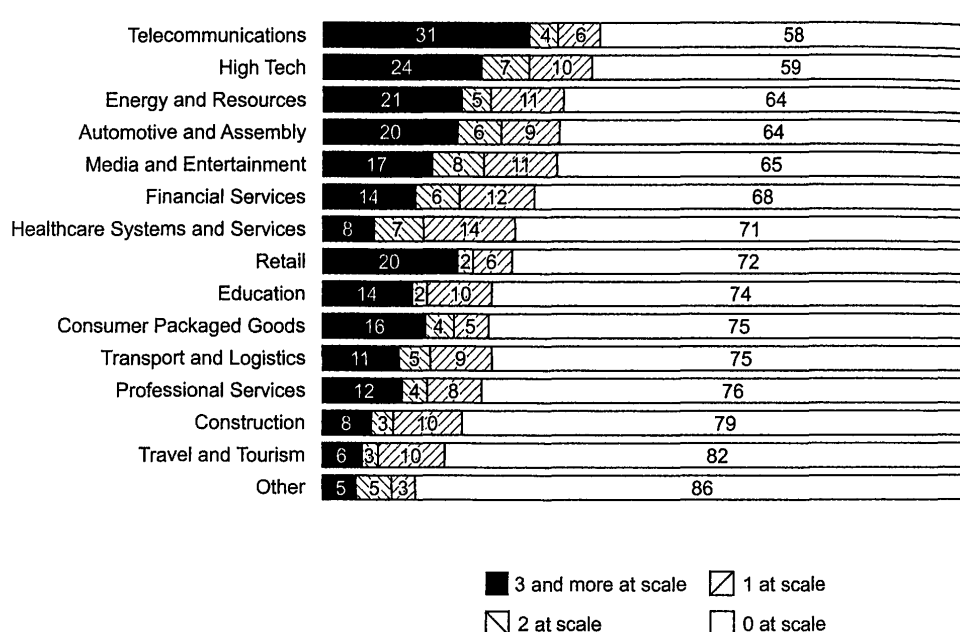


Figure 16 . 4 : 大規模或業務核心部分採用的 AI 相關技術的數量。來源: McKinsey (2017)

Bughin and Hazan (2017) 最近對 3,000 名「AI Aware」C 級高管進行了關於採用準備情況的調查。在這些高管中，有 20% 是「認真採用者」，有 40% 是「實驗者」，還有 28% 的廠商認為他們的廠商「缺乏實施 ML 的技術能力」。McKinsey 確定採用的主要推動力是領導力，



技術能力和數據使用。圖 16.4 區隔了 ML 的採用在各個經濟部門之間的變化。毫不奇怪，電信，科技和能源等產業領先於對建築和旅行等技術要求不高的產業。

### 16.3.1 Machine Learning and Vertical Integration

產業組織的一個關鍵問題是如何將機器學習工具和數據結合起來創造價值。這是否會在廠商內部或內部發生？機器學習用戶會開發自己的機器學習函數還是從供應商那裡購買機器學習解決方案？這是經典的製造與購買問題，這是了解許多現實世界中的產業組織的關鍵。

如前所述，雲端供應商為數據處理和分析提供了集成的硬體和軟體環境。他們還提供對公共和私有數據庫的使用，提供標籤服務，諮詢和其他相關服務，這些服務使一站式購物可以進行數據處理和分析。由雲端提供商提供的專用硬體（例如 GPU 和 TPU）已成為區分提供商服務的關鍵技術。

像往常一樣，標準化與差異化之間存在著張力。雲端提供商正在激烈競爭以提供易於維護的標準化環境。同時，他們希望提供與眾不同的服務。

數據處理和機器學習是在產品速度和性能方面競爭的自然領域。

### 16.3.2 Firm Size and Boundaries

機器學習會增加還是減少最小有效規模？答案取決於固定成本和可變成本之間的關係。如果廠商必須花費大量資金來開發針對其問題的定制解決方案，我們可能會認為固定成本是巨大的，並且廠商規模必須很大才能攤銷這些成本。另一方面，如果廠商可以從雲端供應商那裡購買現成的服務，那麼我們預計固定成本和最小有效規模將很小。

例如，假設一家換油服務廠商希望按名字歡迎回頭客。他們可以使用將車牌號與客戶名稱和服務歷史聯繫起來的數據庫來完成此任務。對於小型提供商而言，編寫軟體來實現此目的將是非常昂貴的，因此只有大型連鎖廠商才能提供此類服務。另一方面，第三方可能會開發智慧手機應用程式，以像徵性的價格提供此服務。這項服務可能會降低最小有效規模。相同的考慮因素也適用於其他小型服務提供商，例如餐館，乾洗店或便利商店。

如今，由於有多家商業服務提供商，因此新的初創廠商能夠外包各種商業流程。就像快餐店提供者可以在一個機構中完善一個模型然後進入全國範圍一樣，商業服務廠商可以一次構建系統並在全球範圍內複製它們。

這裡列出了一家初創廠商如何將十幾個業務流程外包。

- 在 Kickstarter 上資助您的項目。

- 備份來自 Google, Amazon 或 MicroSoft 的雲端計算和網路。
- 使用開源軟體, 例如 Linux, Python, Tensorflow 等。
- 使用 GitHub 管理軟體。
- 成為微型跨國廠商並從國外聘請程式員。
- 設定 Kaggle 機器學習競賽。
- 使用 Skype, Hangouts, Google Docs 等進行團隊交流。
- 將 Nolo 用於法律文件 (廠商, 專利, NDA)。
- 使用 QuickBooks 進行記帳。
- 使用 AdWords, Bing 或 Facebook 進行行銷。
- 使用 ZenDesk 獲得用戶支持。

這只是部分列表。矽谷和 SOMA 的大多數初創廠商都利用了其中的一些業務流程服務。通過選擇標準化的業務流程, 初創廠商可以專注於其核心競爭力, 並在規模擴展時根據需要購買服務。人們將期望由於這些業務流程服務的可用性而獲得更多的進入和更多的創新。

### 16.3.3 Pricing

雲端計算和機器學習的可用性為根據客戶特徵調整價格提供了很多機會。拍賣和其他新穎的定價機制可以輕鬆實施。價格可以輕鬆調整的事實意味著可以實施各種形式的差異定價。但是, 必須記住, 客戶並非無可奈何。他們還可以利用增強的搜索函數。例如, 航空廠商可以採用將購買價格與出發日期聯繫起來的策略。但是可以創建對航空演算法進行逆向工程並建議消費者何時購買的服務 (例如, 參見 Etzioni et al. (2003)。參見 Acquisti and Varian (2005), 了解消費者如何應對嘗試購買商品的理論模型。根據消費者歷史記錄價格以及消費者如何應對此類嘗試。

### 16.3.4 Price Differentiation

傳統上, 價格差異分為三類:

1. 第一級 (個性化),

2. 第二級 (版本: 所有消費者的價格選單相同, 但價格隨數量或品質而變化), 以及
3. 第三級 (基於會員資格的團體定價)。

完全個性化的定價是不現實的, 但是基於消費者細粒度特徵的價格可能是可行的, 因此, 第三級和第一級之間的界限變得有些模糊。Shiller (2013) 和 Dube and Misra (2017) 研究了使用 ML 模型可以提取多少消費者剩餘。

二級價格歧視也可以看作是組成員的定價, 但要認識到組成員和行為的內生性。使用觀測數據進行機器學習在設計此類定價方案方面的幫助有限。但是, 強化學習技術 (例如多臂式吃角子老虎機) 也可能會有所幫助。

根據大多數非經濟學家的觀點, 唯一比價格差異更糟糕的是價格歧視! 但是, 大多數經濟學家認為, 從效率和公平的角度來看, 價格差異通常是有益的。價格差異可以為原本就無法提供服務的市場提供服務, 並且那些不提供服務的市場通常會涉及低收入消費者。

DellaVigna and Gentzkow (2017) 表示: 「我們記錄的統一定價顯著提高了相對於富人而言, 較貧窮家庭所支付的價格。」這種影響可能很大。作者表明: 「收入最低十分位數食品商店的消費者支付的價格比靈活定價高出約 0.7%, 但收入最高十分位數商店的消費者的價格比彈性定價低 9.0% 價錢。」

### 16.3.5 Returns to Scale

至少有三種類型的規模化報酬可能與機器學習相關。

1. 古典的供應方規模收益 (平均成本下降)。
2. 需求方規模報酬 (網路效果)。
3. 邊做邊學 (由於經驗而提高品質或降低成本)。

#### *Supply-Side Marginal Returns*

看起來軟體似乎是供方大規模獲得報酬的範例: 開發軟體的固定成本很大, 而分發軟體的成本卻很小。但是, 如果我們將這種公認的簡單模型與現實世界進行比較, 則存在一個緊迫的問題。

軟體開發不是一次性的操作。隨著時間的推移, 幾乎所有軟體都會得到更新和改進。以手機操作系統為例: 經常每月發布錯誤修復和安全性改進, 再加上每年發布的重大升級。

請注意，這與有形商品有很大的不同確實，汽車中存在一些針對機械問題的錯誤修復，但隨著時間的流逝，汽車的函數或多或少保持不變。Tesla 品牌是一個明顯的例外，該品牌會定期發布新的更新的操作系統。

隨著越來越多的產品啟用網路函數，我們可以預期這種情況會更頻繁地發生。您的電視曾經是一台靜態設備，將能夠學習新的技巧。現在，許多電視都具有語音交互函數，我們可以預期，機器學習將在這一領域繼續發展。這意味著您的電視將變得越來越善於交流，並且可能會更好地識別您對各種內容的偏好。其他設備也是如此它們的函數在銷售時將不再固定，而是隨著時間的推移而發展。

這引起了有關商品和服務之間區別的有趣的經濟問題。當某人購買手機，電視或汽車時，他們不僅在購買靜態商品，還在購買允許他們使用全部服務的設備。反過來，這引發了有關價格和產品設計的一系列問題。

#### *Demand-Side Returns to Scale*

需求方規模經濟或網路效果具有不同的種類。存在直接的網路效果，其中產品或服務對增量採用者的價值取決於其他採用者的總數，而存在間接的網路效果，其中存在兩種或多種類型的互補採用者。用戶更喜歡具有許多應用程式的操作系統，而開發人員更喜歡具有許多用戶的操作系統。

直接的網路效果可能與機器學習系統中使用的編程語言的選擇有關，但是主要語言是開源的。同樣，潛在的用戶可能會更喜歡擁有很多其他用戶的雲端提供商。但是，在我看來，這與許多其他產業沒有什麼不同。由於購買者，維修店，零件和機械師很容易獲得，因此汽車購買者可能更喜歡流行品牌。

在律師和管制機構之間流傳著一個稱為「數據網路效果」的概念。該模型是擁有更多客戶的廠商可以收集更多數據，並使用這些數據來改進其產品。這通常是正確的改善營運的前景使 ML 更具吸引力但這並不是什麼新穎的事。這當然不是網路效果！這實質上是一種供應方面的影響，稱為「邊做邊學」（也稱為「經驗曲線」或「學習曲線」）。經典的展示是 Arrow (1962)。Spiegel and Hendel (2014) 包含了一些最新的引用和引人注目的示例。

#### *Learning by Doing*

邊做邊學通常被建模為一個過程，在此過程中，單位成本隨著累計生產或投資的增加而下降（或品質提高）。粗略的經驗法則是，產量翻倍會導致單位成本下降 10% 到 25%。儘管尚不能確定提高效率的原因，但重要的是，邊做邊學需要廠商的意圖和投資，這在 Stiglitz and Greenwald (2014) 中有所描述。

這通過邊做邊學來區別於通常被認為是或多或少自動的需求方或供應方網路效果。這也不

是真的 在存在網路效果的情況下，整本書都涉及策略行為。但是邊做邊學與所謂的「數據網路效果」之間存在重要區別。廠商可以擁有大量數據，但是如果對數據不執行任何操作，則將不會產生任何價值。

以我的經驗，問題不是缺乏資源而是缺乏技能。一家擁有數據但無人分析的廠商無法利用這些數據。如果內部沒有現有的專業知識，則很難明智地選擇所需的技能以及如何找到和僱用具有這些技能的人員。僱用優秀人才一直是競爭優勢的關鍵問題。但是由於數據的廣泛可用性是相對較新的，所以這個問題特別嚴重。汽車廠商可以僱用知道如何製造汽車的人，因為這是他們核心能力的一部分。他們可能沒有，也可能沒有足夠的內部專業知識來僱用優秀的數據科學家，這就是為什麼我們可以期望看到這種異質性，因為這種新技能滲透到勞動力市場。Bessen (2016, 2017) 就此問題撰寫了感性的文章。

### 16.3.6 Algorithmic Collusion

幾十年來，眾所周知，重複賽局中存在許多均衡。該領域的主要結果是所謂的「民間定理」，即在重複賽局中幾乎所有結果都可以作為均衡來實現。有關此結果的各種表述，請參見 Fudenberg (1992) 和 Pierce (1992) 的調查。

寡占群體的互動可以看作是重複的賽局，在這種情況下，應特別注意合謀結果。有非常簡單的策略可用於促進合謀。

*Rapid Response Equilibrium.*—例如，考慮一下在街對面的兩個加油站的經典示例，它們可以迅速改變價格並為固定的消費者群體提供服務。最初，它們的定價都高於邊際成本。如果其中一個降價一美分，那麼另一個將很快與該價格匹配。在這種情況下，兩個加油站的情況都會惡化，因為它們的售價較低。因此，降價沒有任何報酬，高價佔據了上風。如 Varian (2000) 中所述，這種策略可能已用於線上競爭中。Borenstein (1997) 在機票價格中記錄了相關行為。

*Repeated Prisoner's Dilemma.*—在 1980 年代初期，Robert Axelrod (1984) 舉辦了一次囚犯困境競賽。研究人員提出了反覆對戰的演算法策略。獲勝者大都是 Anatol Rapoport 提出的一種簡單的策略，稱為「針鋒相對」。在這種策略中，雙方開始合作（收取高價）。如果任一玩家有缺陷（降低其價格），則另一玩家匹配。然後，Axelrod 建立了一個錦標賽，根據比賽中的收益複製策略。他發現表現最佳的策略與針鋒相對。這表明人工代理可能會學會在經典的雙頭獨占賽局中扮演合作策略。

*NASDAQ Price Quotes.*—在 1990 年代初期，NASDAQ 的報價是八分之一美元而不是美分。因此，如果出價為八分之三，而要價為八分之二，則交易發生，買方支付八分之三，賣方支付八分之二。買入價和賣出價之間的區別是「內部價差」，它補償了交易者承擔風險並維持參與市場

所需的資本。請注意，內部價差越大，做市商對做市商的補償就越大。

在 1990 年代中期，兩位經濟學家 William Christie 和 Paul Schultz 檢查了 NASDAQ 排名前 70 位的上市廠商的交易，令他們驚訝的是，實際上沒有以奇數價格進行交易。作者得出的結論是：「我們的結果很可能反映了做市商之間的諒解或隱含協議，以避免在報價這些股票時使用八分之一的價格分數」(Christie and Schultz 1995, 203)。

司法部 (DOJ) 展開了隨後的調查，最終以 10.1 億美元的罰款得以了結，當時這是有史以來在反托拉斯案件中支付的最大罰款。

正如這些示例所說明的，在重複交互的上下文中，隱式 (或可能是顯式) 合作似乎有可能發生，Axelrod 稱之為「合作進化」。

近來，在「演算法勾結」的背景下，這類問題再次出現。2017 年 6 月，經濟合作與發展組織 (OECD) 舉行了一次關於演算法和合謀的圓桌會議，作為其數位經濟競爭工作的一部分。見 OECD (2017) 的背景文件以及 Ezrachi and Stucke (2017) 對圓桌會議的代表性貢獻。

在這種情況下，出現了許多有趣的研究問題。民間定理表明，合謀結果可以是重複賽局的均衡，但沒有描述導致這種結果的特定演算法。眾所周知，非常簡單的演算法 (例如具有少量狀態的有限自動機) 無法發現所有均衡點 (參見 Rubinstein 1986)。

有類似拍賣的機制可以用來近似獨占結果。參見 Segal (2003) 的示例。但是，在寡占獨占的背景下，我還沒有看到類似的機制。

## 16.4 Structure of ML-Provision Industries

到目前為止，我們已經研究了使用機器學習的產業，但是關注提供機器學習的廠商也很有趣。

如上所述，機器學習供應商可能會提供幾種相關服務。立即提出的一個問題是，在提供商之間進行轉換將有多麼容易。專門開發的技術 (例如容器) 使將應用程式從一個雲端提供商移植到另一雲端提供商變得容易。諸如 docker 和 kubemetes 之類的開源實現已經很容易獲得。鎖定對於中小型應用程式來說不是問題，但當然，可能存在涉及定制應用程式的大型和複雜應用程式的問題。

由於易於在晶片，主板，機架或數據中心本身的級別上複製硬體安裝，因此計算機硬體也至少顯示出固定的規模報酬。固定報酬的經典複製論點適用於此，因為增加容量的基本方法是複製以前所做的事情：向處理器添加更多核心，向機架添加更多板，向數據中心添加更多機架以及構建更多數據中心。

前面我曾建議，對於大多數用戶而言，雲端計算比從頭開始構建數據中心更具成本效益。有

趣的是，需要大量數據處理能力的廠商已經能夠複製其現有基礎架構，並將額外的容量出售給其他較小的實體。結果是產業結構與經濟學家的想像有些不同。汽車廠商會增加產能，然後將其出售給其他廠商嗎？這並非聞所未聞，但很少見。同樣，正是這種計算的通用性質才使該模型成為可能。

### 16.4.1 Pricing of ML Services

與任何其他基於訊息的產業一樣，軟體生產成本高昂，而複製成本低廉。如上所述，由於易於在晶片，主機板，機架或數據中心本身的水準上複製硬體安裝，計算機硬體還表現出至少固定的規模報酬。

如果服務變得高度標準化，那麼很容易陷入類似於 Bertrand 的降價。即使在早期，機器價格似乎仍然具有很強的競爭力。例如，在所有主要的雲端提供商處，圖像識別服務的成本約為每幅圖像十分之一美分。大概，我們將看到供應商試圖在速度和函數維度上與眾不同。那些可以提供更好服務的廠商可能能夠提供溢價，只要用戶願意為溢價服務付費。但是，當前的速度和準確性非常高，還不清楚用戶如何看待這些尺寸的進一步改進。

## 16.5 Policy Questions

我們已經討論了涉及數據所有權，數據使用，差別定價，規模收益和演算法勾結的問題，所有這些問題都具有重要的政策方面。剩下的主要策略領域是安全和隱私。我首先談談安全性。

### 16.5.1 Security

關於安全性出現的一個重要問題是廠商在這方面是否有適當的誘因措施。在經典文章中，Anderson (1993) 比較了美國和英國關於自動櫃員機 (ATM) 的政策。在美國，除非銀行可以證明他們是錯誤的，否則用戶是正確的 而在英國，除非用戶可以證明他們是錯誤的，否則銀行是正確的。這項責任分配的結果是，美國的銀行投資了安全措施，例如安全攝像機，而英國的銀行卻沒有理會這種基本的預防措施。

這個產業表明責任分配對於創建適當的證券投資誘因機制有多麼重要。侵權法的法律和經濟學分析有助於理解不同責任分配的含義以及最佳分配的外觀。

出現的一項原則是「適當護理」標準的原則。如果廠商遵循某些標準程式，例如在發布安全修補程式後的幾天內安裝安全修補程式，實施兩因素身份驗證，對員工進行安全實踐教育等，那麼他們對與安全事件相關的費用承擔責任是安全的。

但是，適當護理標準從何而來？一種可能性來自政府，特別是來自軍事或執法實踐。The Orange Book 及其後繼書 the Common Criteria standard 就是很好的例子。另一可能性是，保險代理機構向實施良好做法安全性的各方提供保險。就像保險廠商可能需要灑水系統來提供火災保險一樣，網路保險也可能僅提供給從事最佳實踐的廠商（有關更多討論，請參見 Varian 2000）。

該模型是解決該問題的一種有吸引力的方法。但是，我們知道，涉及保險的許多問題，例如逆選擇和道德風險，都需要解決。有關此領域的更多工作，請參閱 Workshop on the Economics of Information Security 的檔案有關概述，請參閱 Anderson (2017)。

## 16.5.2 Privacy

隱私政策是一個廣闊而又廣闊的領域。Acquisti, Taylor and Wagman (2016) 對經濟文獻進行了全面回顧。

機器學習領域出現了一些政策問題。例如，廠商是否有適當的動機來提供適當級別的隱私？隱私與經濟表現之間的權衡是什麼？眾所周知，隱私法規可能會限制 ML 供應商合併來自多個來源的數據的能力，並且跨廠商邊界的數據傳輸和 或數據銷售可能會受到限制。在該領域頒布法規的趨勢會導致意想不到的後果。一個例子是 the Health Insurance Portability and Accountability Act of 1996 (健康保險便利和責任法案)，通常稱為 HIPAA。該立法的初衷是通過建立醫療記錄保存標準來刺激保險廠商之間的競爭。但是，許多研究人員認為，這對醫學研究的數量和品質產生了重大的負向影響。

## 16.5.3 Explanations

歐洲管制機構正在研究「解釋權」的概念。假設將有關消費者的訊息輸入到模型中，以預測他或她是否會償還貸款。如果消費者被拒絕貸款，他們為什麼要被「解釋」？如果是這樣，什麼可以解釋？組織是否可以保守預測模型的秘密，因為它可以被操縱嗎？一個明顯的例子是判別清單函數。即 IRS (Internal Revenue Service, 美國國家稅務局) 用於觸發審核的 DIF 函數 (Discriminant Inventory Function System)。對 DIF 函數進行反向工程是否合法？有關 DIF 函數上的鏈接的集合，請參閱 CAVQM (2011)。

我們能不能比一個人對 ML 模型有更多的要求？假設我們給您看了一張照片，並且您正確地將其識別為配偶的照片。現在我們問：「你怎麼知道？」最好的答案可能是「因為我看過很多照片，我知道這些照片是我配偶的照片，而且這張照片看起來很像那些照片！」來自計算機的解釋是否令人滿意？



## 16.6 Summary

本章僅介紹了 AI 和 ML 如何影響產業結構的概況。技術正在迅速發展，目前的主要瓶頸是可以實施這些機器學習系統的分析師。鑑於該領域大學課程的廣泛普及以及線上教程的豐富性，我們預計這一瓶頸將在未來幾年內得到緩解。

### Comment Judith Chevalier

Varian 很好地概述了由於採用機器學習和人工智能而引起的產業組織問題。其中許多問題都可能對競爭政策產生影響。例如，利用 AI 技術可能會增加或減少規模經濟，從而潛在地導致市場勢力。數據所有權，如果對特定產業的競爭至關重要，則可能會造成進入壁壘。演算法勾結的潛力顯然會引起反托拉斯執法的關注。在這裡，我簡要介紹一下這些問題之一，即數據所有權，並重點介紹一些潛在的反托拉斯政策應對措施。雖然我在這裡將數據所有權作為進入壁壘的重點，但我所討論的一些政策取捨與 Varian 強調的其他潛在市場結構變化密切相關。

人工智能和機器學習過程經常使用原始數據作為輸入。正如 Varian 指出的那樣，目前尚不清楚數據是否符合我們通常的預期，即稀缺資產或資源最終將面臨規模收益遞減的預期。但是，人們當然可以想像這樣一種情況：對數據主體的專有所有權將為市場老牌廠商帶來幾乎不可逾越的優勢。儘管對於稀缺資產的獲取會產生進入壁壘的擔憂可能相對較新，因為它適用於數據，但根本的基本經濟問題並不新鮮。長期以來，所有司法管轄區的反托拉斯當局都對針對稀有資產所有權構成進入壁壘的廠商的最佳政策進行了角力。在美國，對該問題的分析至少可以追溯到 *United States v. Terminal Railroad Association* (224 US 383 (1912)) 一案，在該案中，鐵路財團拒絕競爭對手進入穿越聖路易斯河的唯一鐵路橋樑在這種情況及以後的情況中，法院偶爾會明確規定有義務與具有市場支配力的廠商交易，該市場支配權控制了對競爭至關重要的資產（或設施）的獲取，而競爭對手則無法複製該資產（或設施）。然而，確定獨佔者在與競爭對手競爭時負有肯定義務的確切情況仍然是反托拉斯法尚待解決的領域。

原則上，這種反托拉斯必不可少的設施原則可以應用於數據所有權。的確，儘管 Varian 在補救措施問題上保持沉默，但美國最近的法律文獻對將基本設施學說應用到數據上表現出了一定的熱情（例如參見 Meadows 2015 Abrahamson 2014）。此外，歐洲反托拉斯當局已經開始闡明控制大數據的原則，這些原則提出了基本的設施理論。例如，歐盟競爭事務專員 Margrethe Vestager (2016) 最近在一次演講中表示：「的確，我們不應該對每個擁有有價值數據集的廠商都抱有懷疑。但是，我們確實需要密切關注關於廠商是否控制著獨一無二的數據，這是其他人無法控制的，並可以利用這些數據將其競爭對手拒之門外。在演講中，她強調了 2014 年的一樁案

例，法國競爭管理局要求法國能源生產商 GDF Suez 與產業競爭對手分享客戶名單。

儘管在某些方面表現出熱情，但將基本設施學說應用於數據共享卻產生了重要的權衡和重要的實際問題。我從權衡開始。在評估創新產業的反托拉斯政策時，重要的是要認識到，新技術給消費者帶來的好處不僅來自以具有競爭力的價格獲得商品和服務，而且還源於創新產生的新的和改進的產品和服務。因此，應該不僅根據反獨佔政策對價格和產出的影響，而且還應根據其對創新速度的影響來評估反托拉斯政策。實際上，在高科技產業中，這些動態效率考慮因素似乎使靜態效率考慮因素相形見拙。在將基本設施學說應用於數據的情況下，折衷方案很多，方向也不明確。

人們經常引用對基本設施學說的批評是，設立事後分享義務首先會削弱對基本設施投資的動機（例如，見 Pate 2006）。在這種情況下，創建事後共享數據的職責可能會減少對數據創建進行投資的現有動機，從而減慢了創新的步伐。但是，總的誘因折衷並不是那麼簡單。在新進入者是潛在創新的重要來源的情況下，現有廠商的排他性行為會降低新進入者對研發投資的動機，這會減慢創新的步伐。也就是說，就數據而言，如果特定數據是 AI 創新的必要補充，那麼任職者對數據的專有所有權可能會減慢參與者的創新步伐。反托拉斯執法對創新步伐的影響問題仍是一個新興的研究領域，但在理論上已在例如 Segal and Whinston (2007) 中進行了探討。因此，總而言之，雖然從事後靜態效率的角度出發，將基本設施學說廣泛應用於專有數據可能很誘人，但仍應謹慎對待事前誘因措施。

除了已經討論過的權衡之外，將基本設施理論應用於數據共享的任何應用都還意味著許多實際考慮。如同在任何基本設施方案中一樣，法院或反托拉斯機構一旦確立了交易義務，就必須明確規定貿易條款。顯然，由於缺乏條款的明確規定，在位者可以通過建立對數據的任何潛在競爭對手都沒有吸引力的交易條款來抗拒拒絕交易。鑑於市場條件不斷變化，對貿易條件的持續管制將不可避免。當然，在某些情況下，美國法院已成為法院強制執行交易義務的廠商交易的持續管制者。音樂許可廠商 ASCAP 和 BMI 的契約的持續監督是有責任履行交易的很好的例子，導致法院事實上進行了管制。但是，建立這種持續的管制結構會給管制實體和受管制廠商帶來成本。基本設施不是快速解決方案。

最後，雖然基本設施原則可能不一定總是解決所有權已集中的數據的最佳工具，但在合併分析中應考慮合併產生重要集中數據的潛力，就像合併分析認為合併可能充分集中的潛力一樣生產能力的其他要素。

顯然，在實施反托拉斯解決方案時，在權衡關鍵數據專有權可能帶來的問題時需要進行重要的權衡。這至少提出了其他一些可供探索的公共政策途徑。例如，考慮到數據的公共物品性質，在某些情況下值得在數據創建和對這樣創建的數據的公共所有權上進行公共投資，特別是在私

人創建此類數據會引起反托拉斯關注的情況下。

譯 者: 賴孚權

單 位: 中央研究院人文社會科學研究中心研究員

打 字: 任柏勳

檔案名稱: Varian2019\_Ch16\_inChinese\_2020617\_fcn.ctx

完稿日期: June 22, 2020 1:44pm