



指南

計劃：分析

文件編號：T147



NUCLEUS
RESEARCH

指南

在 AWS 上的深度學習

分析

Daniel Elman

Nucleus Research, Inc.

100 State Street, Boston, MA, 02109

+1 (617) 720-2000

www.nucleusresearch.com

©2019 Nucleus Research, Inc.

基本論點

深度學習仍是人工智慧 (AI) 領域中最熱門的主題之一，並且正迅速從學術界和理論領域轉向可操作的增值工作負載。資料基礎架構和運算能力的進步，以及新型神經網路的出現，讓深度學習成為現代企業使用的可行方法，而一年前，這仍只是一個空想。與 Amazon Web Services (AWS) 這類雲端供應商合作，以及充分利用 Amazon SageMaker 這類雲端託管的機器學習服務，對於希望將深度學習專案從概念加速運用至生產的公司來說十分重要。過去，基礎架構並不先進，機器學習服務亦不成熟或尚未可行，財力雄厚的大型公司及開發人員團隊著手研究並實現這些 AI 功能時，缺乏預算和專業知識的企業無法從頭開始在內部發展，只能作為旁觀者。

為了解現今深度學習的採用和使用情況，以及自去年份的報告 (Nucleus Research s180 – *Guidebook: TensorFlow on AWS* – 2018 年 11 月) 發布以來的變化，Nucleus 進行各項訪問，並分析了 316 個獨特專案的經驗。兩年來，生產中的深度學習專案數目兩度增加了兩倍以上。我們發現 96% 的深度學習是在雲端環境執行，其中 TensorFlow 是最受歡迎的框架，74% 的深度學習專案中使用了 TensorFlow。43% 的專案使用了 PyTorch (請注意，大多數專案利用多個框架；還使用了 MXNet、Keras 和 Caffe2，幾乎總是與 TensorFlow、PyTorch 或兩者結合使用)，相較於去年，採用率顯著增加。在 316 個專案中，只有 9% 使用單一框架建構。最值得注意的是，在雲端託管的深度學習專案中，有 89% 在 AWS 上執行；主要原因是 AWS 上有廣泛的框架選擇，以及其對現有和新服務的持續投資。

我們還發現，以雲端為主的 TensorFlow 工作負載中，有 85% 在 AWS 上執行，而以雲端為主的 PyTorch 專案中，有 83% 在 AWS 上執行。去年，大約三分之一的受訪客戶正在考慮或使用 SageMaker，這是 Amazon 受管服務，用於大規模建構、訓練、部署和協調深度學習模型。在今年的受訪使用者中，有 63% 的 Amazon 客戶已開始使用 SageMaker。

形勢

機器學習包含電腦透過分析資料「學習」經驗，無須人員參與的技術。深度學習是機器學習的子集；在機器學習中，電腦獲得包含指定功能集的資料以進行分析，而在深度學習中，電腦獲得非結構化資料 (例如文字、音訊或視訊)，並自行確定哪些功能與分析相關。簡而言之，電腦具有成對的範例輸入和相應的輸出，並且能夠回溯工作，以尋找將輸入轉換為輸出所需的操作。

現代深度學習模型需要大量的運算和儲存，大多數企業無法自行建構這些系統。因此，我們在研究過程中發現，企業絕大多數都希望在深度學習專案中充分利用雲端。這種方法能讓企業去購買專案所需的資料儲存量和運算能力，而不必在內部採購、設定和維護基礎架構，長期下來可節省大量成本。

由於 AI 的大肆宣傳和潛力，使用群體不斷擴大，且技術迅速發展。新的工具和方法也不斷推出。因此，使用者能夠以最大的靈活性在平台上進行開發，是雲端機器學習市場能長期成功的關鍵。簡而言之，雲端平台需要支援現在和未來使用的各種工具和開發框架，並具有必要的安全性和可用性，以遵守資料處理和隱私權法規。

這已是連續第三年進行這項研究，並且在這三年中，我們看到了模型功能、運算能力和開發人員工具的翻轉型變化，帶來了新的且令人驚喜的結果。在第一年，很難找到超越初步開發和具有深度學習概念驗證專案的企業。在 2018 年，我們看到了大幅的進展，其中 14% 的專案被歸類為生產中的、處理即時資料的專案。今年也有廣大進步，從 20 人的初創公司到《財富》100 強全球企業，各種規模的企業均部署深度學習來生產，其中投入生產的專案佔 38%。採訪中其它的討論事項：

- 深度學習專案背後的目標和動機
- 部署策略及相關優勢
- 正在使用的開發框架、方法和其他工具
- 不同模型和框架的相對優勢和劣勢
- 專案團隊中的人數和各自的角色

Nucleus 共對 32 位深度學習專家進行了深度訪問，其中許多人同時負責多個專案，代表了 316 個獨特專案。

雲端中的深度學習

生產規模的深度學習工作負載涉及處理成千上萬的範例資料，以訓練模型。這在運算方面成本非常昂貴，對於影像或視訊等複雜的輸入資料尤其如此，大多數企業無力負擔建構和維護搭載平行 CPU 或 GPU，來執行運算的高效能運算系統。因此期望依靠雲端來存取其所需的資源和基礎架構。今年，我們發現 96% 的深度學習專案都在雲端環境執行，這與去年的研究結果相

吻合，隨著 2018 年的 177 個專案增加至 2019 年的 316 個，仍然代表雲端執行機器學習擁有強大的客戶潛力。在生產中的即時資料工作負載中，有 98% 在雲端執行。對於尚未完全在雲端部署的企業，常見的部署策略包括，在移轉至雲端進行生產之前，針對現場的硬體開發小型模型。

雲端的優勢

有 96% 的深度學習專案在雲端執行，顯然客戶體認到了它的價值。我們請受訪者出提在雲端執行深度學習的主要優勢。他們的回覆主要環繞三個關鍵主題：

- 因減免硬體、人員和能源成本而減少了開銷。這是每個受訪者提到的最常見的回答。深度學習需要大量的運算；建立和維護可大規模進行深度學習的硬體系統，需要專門的 IT 專業人員；實體上執行硬體以訓練深度學習模型會耗費數千小時的 CPU 和 GPU 時間，單電力成本往往十分高昂。藉由雲端，使用者按照其用量付費，而無須付其它相關費用。
- 能夠在不同團隊中合作和工作。無論實體位置如何，所有許可使用者都可存取部署在雲端的模型。這加快了模型開發的速度，尤其是在越來越普遍的雲端團隊之間。
- 能夠充分利用補充平台功能和工具。安全性和可用性是最常提到的方面。憑藉客戶和雲端供應商的安全投資，正確設定的雲端系統將大受裨益。此外，能夠在本地資料中心執行模型，以符合 GDPR 等資料保護法也很重要。除了儲存和運算，雲端供應商還提供工具和平台功能，來改善開發人員體驗。Amazon SageMaker 這類工具透過全受管端對端機器學習工作流程，為雲端客戶帶來了巨大價值，包括清理資料，以及訓練、建構和部署模型。

名副其實的深度學習

去年，在 14% 的生產專案中，機器學習向前邁出了一大步。我們在 2019 年看到了類似的大幅進展，在 316 個深度學習專案中，有 38% 被歸類為生產中專案。在生產中的深度學習專案中，有 89% 在 AWS 上執行。生產中有 76% 的專案使用 TensorFlow，而 28% 的專案使用 PyTorch。Keras 和 Apache MXNet 也出現在生產環境中，因為大多數專案的元件都使用多個框架建構。只有 9% 的專案僅使用一個框架建構。在企業體認到深度學習和其它 AI 功能已經達到成熟的水準，能夠提供認可的商業價值時，不約而同地搶著執行較容實現的目標，或經

過充分示範並可快速實作的簡化使用案例。常見範例包括網站和應用程式的語音界面，以及線上購物網站的推薦引擎。

許多企業仍在探索更複雜的深度學習應用程式，但許多仍在測試中。今年的研究結果顯示，企業從去年開始擴大了對深度學習的總體投資，在更複雜、多年期且雄心勃勃的專案上繼續進行，同時增加了快速實作的增值應用程式，例如推薦、聊天機器人中的情感分析，以及語音界面，以跟上市場趨勢，並證明深度學習在實際使用中的持續價值和可行性。

AWS 上的深度學習

Nucleus 發現，選擇 AWS 的主要原因 (平台功能的廣度、與 Amazon 的關係，以及 AWS 繼續投資於深度學習服務) 自去年以來一直保持不變。

AMAZON 功能的廣度

Amazon 支援端對端的深度學習程序，因此不須將來自不同供應商的同類最佳元件組合在一起。客戶能夠在平台上儲存資料，建構和部署模型，以及建立充分利用模型輸出的應用程式。客戶可以靈活地選擇針對其工作負載最佳化的專用硬體；例如，他們可以存取針對深度學習最佳化的強大 GPU 執行個體，如 Amazon EC2 P3 和 G4。Amazon 在全球範圍設有區域資料中心，客戶可以視需要將資料和營運在地化，並遵守區域資料共用法規。平台級安全性為資料和應用程式新增另一層保護。平台上的其它整合工具和服務 (如容器化服務) 非常適合大型專案。使用者表示：

- 「其他雲端供應商無法與 AWS 平台的成熟度和完整性匹敵。透過與 AWS 合作，我們獲得的不僅僅是技術；他們擁有 S3 和 CloudFormation 等專業知識和附加服務，可協助我們將技術專案從概念引入生產。」
- 「我們一直在評估其他雲端供應商，但迄今為止，沒有人能與 AWS 相提並論。其他人都無法在雲端提供與之比擬的價值，更不用說一流的平台安全性和可用性。」

使用者檔案 – 生物科技公司

一間生物科技公司正使用深度學習來開發生產級分類器，以根據基因資料預測癌症。由於它仍處於研究階段，正在建構模型以平行測試不同的假設，因此以靈活的方式快速實作和測試新想法非常重要。

有四種平行測試的方法。其中兩個是在 TensorFlow 上建構，一個在 PyTorch 上建構，另一個則使用完全自訂的代碼開發。框架的選擇具有策略性，針對預先建立的估算器和損失函數，兩個專案選擇 TensorFlow；之所以選擇 PyTorch，是因為該方法受到基於 PyTorch 的學術界專案的啟發。

而公司選擇 AWS 來實作其深度學習專案，則是出於以下多種原因。Amazon 針對不同的使用案例，提供了最適合的運算選項。基因資料的深度學習需要大量的預處理，因此該公司需要針對記憶體最佳化的機器來預處理訓練資料，以及需要 GPU 加速的機器來訓練模型。由於要使用 PB 級的訓練基因資料，雲端合作夥伴是否能支援如此規模的專案十分重要。由於專案採用不同的框架，因此雲端供應商必須在不進行大量設定的情況下，在其基礎架構上以第一等級支援多個框架，這一點很重要。Amazon 與之提供的效能相得益彰，即使是 PB 級，其支援的機器極具靈活性，並針對所有框架最佳化，才能表現優異。

與 AMAZON 的關係

受訪客戶還提到他們與 AWS 的關係是業務決策背後的驅動因素。許多客戶已經在業務的其它領域使用 AWS，並增加了投資在深度學習上。深度學習需要大量資料，而企業則信任 AWS 處理其資料。此外，Amazon 還為客戶提供了完成專案所需的技術和最佳實作方式，而無須限制在任何特定的解決方案中。客戶很快就強調了平台的靈活性，他們可以使用認為最舒適的框架和資料庫，而不必擔心是否與 AWS 技術不相容。一位客戶提到 Amazon 致力於此類研究，證明其承諾實現將客戶價值最大化。其他使用者表示：

- 「我們將整個業務都建立在 AWS 之上，若轉換供應商代價會很高。我們的 AWS [架構] 與我們一起成長，並且每一步都獲得許多支持。我們確實感覺到他們投入了資金來協助我們取得成功，這對任何長期合作夥伴關係都十分重要。」
- 「藉助 Amazon，我們可以針對我們的機器學習專案，存取相關的客戶使用案例和行業專業知識。支援團隊對我們的工作很有幫助，他們能夠回答有關框架、模型選擇和基礎架構要求的所有問題。」

使用者檔案 – 數位媒體和手機遊戲公司

一間設計和銷售手機遊戲的數位媒體公司使用深度學習，來平衡遊戲難度與預計產生的收入。透過使用深度強化學習，訓練機器人玩每一個遊戲來實現這一點。隨後可監控機器人的表現，以估算每個級別的難度，並用於預測每個級別會有多少使用者退出。當然，若使用者停止玩遊戲，該遊戲就不會為供應商產生廣告收入，因此公司會透過此程序來平衡遊戲難度和預計收入。

自 2008 年以來，AWS 一直是該公司的雲端供應商，因此使用 AWS 進行深度學習不費吹灰之力。這兩個專案都是建構在 PyTorch 上，PyTorch 是深度強化學習的行業標準，並且目前的許多研究都隨 PyTorch 實作而發佈。

該公司使用 SageMaker 部署，負責預測玩家消耗到生產中的模型。之所以選擇它，是因為它為專案負責人提供了該模型的「鳥瞰圖」，並為專案負責人提供了一個集中的位置，來檢視並控制部署中的所有模型。透過結合 SageMaker 這類技術與 Amazon 提供的靈活運算資源，受訪專家表示：「要說服我們離開 AWS，需要一個令人信服的商業案例，到目前為止，尚無其他雲端供應商能夠向我們提供機器學習特定工具，以及實現 [更大價值] 的儲存和運算。」

AWS 在機器學習方面的投資

客戶知道 Amazon 正在開發和使用自己的深度學習技術。深度學習專家提到，對文件、框架支援和 Amazon SageMaker 等雲端服務的持續改進，是選擇 AWS 而非其他雲端供應商的主要因素。SageMaker 於 2017 年推出；這是一項全受管雲端服務，涵蓋了整個機器學習工作流程，包括建構、訓練和部署機器學習模型。

隨著開發人員意識到如何降低複雜性並加快模型部署，SageMaker 的採用率正在迅速增長。去年，約三分之一的受訪者表示正在使用或摸索 SageMaker 的使用，以自動化其深度學習專案的各個方面；今年這一數據幾乎翻倍，有 63% 的客戶正在使用或考慮使用 SageMaker。在本研究過程中，我們發現有客戶將自己的 TensorFlow 部署透過 SageMaker 移轉至 AWS 受管服務，以及完全使用 SageMaker 建構其系統的客戶。使用者表示：

- 「我們無須採購專用軟體，即可執行大型深度學習專案。若沒有 AWS，特別是 SageMaker，我們需要購買硬體，在本地訓練模型，然後在內部伺服器上儲存和託管模型，以便在需要進行預測時可以取用。剛開始，這本可能要花幾週的時間，並且會增加大量的硬體、電力和人員成本，更不用說在內部建構功能耗費的所有時間，而這些功能在 SageMaker 已預先建構。」

- 「我們專注於為客戶實作對話式機器學習，而 SageMaker 精簡了模型的建構。平均而言，我們在指定的一週內，同時為客戶進行五個專案，自去年開始使用 SageMaker 以來，我們發現，透過自動化分散式模型訓練，每個專案可節省約兩個小時。」

使用者檔案 – 企業軟體公司

一間主要為銷售和服務團隊生產應用程式的全球企業軟體公司採用 Amazon SageMaker，來管理其 TensorFlow 部署。其深度學習工作主要圍繞情感分析和對客戶互動進行分類，以便有效地了解不同類型的外展活動，如何影響客戶流失或再次購買的可能性。

作為一間大型技術公司，相較於整個市場，其在深度學習方面一直處於領先地位。在使用 SageMaker 之前，該公司客製建構了大部分以 TensorFlow 為主的深度學習基礎架構。在今年年初，它決定將自我管理的 TensorFlow 部署移轉至 SageMaker，在其中將該部署作為服務進行管理。雖然工作正在進行中且尚未完全完成，但到目前為止，該組織已經能夠重新指派 3 個 FTE，主要負責管理 TensorFlow 生態系統。此外，由於 SageMaker 自動在多個 CPU 或 GPU 中平行自動散發運算負載，因此訓練模型的速度大大提高。該公司報告稱，在自我管理的環境中，使用 SageMaker 部署新模型所需的時間不到 50%。

使用者檔案 – 應用程式部署公司

一間專精於開發可在智慧揚聲器 Amazon Alexa 上暢玩語音整合遊戲的應用程式開發公司，完全在 Amazon SageMaker 上建構深度學習專案，透過推薦遊戲以保持較高的使用者參與度。系統利用之前在系統上玩過的遊戲資料，來預測性地向使用者推薦其他類似遊戲。

該公司將其業務建立在 AWS 平台，因此選擇充分利用 SageMaker 與 AWS 上的現有架構 (尤其是 Amazon S3 和 AWS Lambda) 進行原生整合，以存取儲存的資料和無伺服器運算。此外，由於該公司已在 AWS 上，因此使用者權限和開發程序已規範化，從而讓他們能夠避免重複工作。憑藉平行化的模型訓練和控制級 UI，訓練和評估比手動方式要快得多。客戶估計，相較於手動管理系統，使用 SageMaker 可讓模型訓練提速三倍，模型部署提速四倍。

結論

去年，隨著機器學習教育、運算技術、雲端基礎架構和適當的資料集的發展，深度學習在商業上變得可行，深度學習專案走出課堂，商業化走入會議室。今年，這一趨勢繼續發展，越來越多的公司希望大規模實作深度學習來解決實際問題。兩年來，生產中的深度學習專案數目兩度增加了兩倍以上，主要是因為與去年相同的關鍵因素。

- **AWS** 等供應商提高了雲端基礎架構和服務的可用性，以支援資料密集型、運算密集型程序 (如深度學習)。
- 憑藉改進的技術、網路架構和資料集，神經網路變得更準確、功能更強，從而使深度學習達到了先進水準。
- 繼續在社群中進行投資，以透過線上論壇和文件、開放原始碼庫和框架，以及雲端產品，例如針對機器學習最佳化的預建構模型和專用硬體，來與其他深度學習研究人員分享經驗。

隨著摸索深度學習的初始成本下降，我們看到越來越多的企業希望加入競爭。為此，最有效的策略不是尋求閉門造車，而是與擁有基礎架構、專業知識和其他服務的雲端供應商合作，以將深度學習從概念轉化為現實。根據我們的分析，我們發現 **Amazon** 擁有最成熟、最完善企業雲端技術供應商的聲譽，加上它在機器學習服務和平台靈活性方面的特定領域投資，可支援客戶選擇網路架構、開發框架或資料來源，使其成為深度學習專業人員的首選雲端平台。