

【與三星研究院共築未來③】中國三星電子研發中心：

透過機器學習基礎研究 洞察革命性創新技術

於此系列報導中，三星新聞中心專訪來自三星全球研發中心的技術專家，深入了解其工作內容及創新如何造福消費者。



第三位專家是 2020 年加入中國北京三星電子研究院（SRC-B）人工智慧實驗室的主任工程師 Bin Dai，負責網路壓縮與終端裝置上的模型設計與研究。以下將介紹他與團隊的重大研究突破。



系列報導《Q&A》

問：AI 為現今相當重要的先進技術，也持續取得重大突破。
而「機器學習」該如何促成相關創新？

- Lukasz Slabinski (波蘭三星電子研發中心)



答：「機器學習」是將AI新技術應用至日常生活的關鍵推手。
我們 AI 實驗室負責製作模型，將先進演算法導入行動裝置。

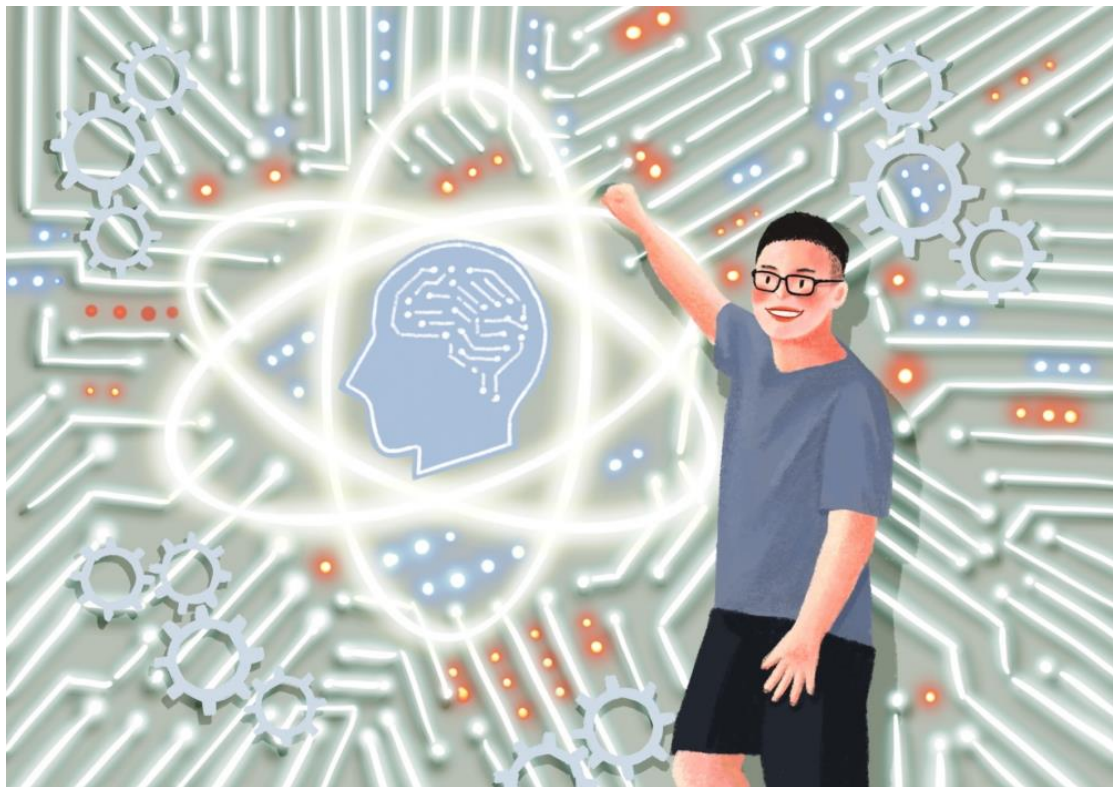
- Bin Dai (中國北京三星電子研發中心)

問：自然語言處理（NLP）和聲學智慧等 AI 技術是目前的前瞻研究領域，且持續取得新進展。與機器學習有關的核心研究是如何促成 AI 技術的創新？

歸功於機器學習，各類技術才得以造福用戶。於 AI 應用領域當中，目前進展最順利的為電腦視覺和語音辨識。現有的先進 AI 演算法需要使用大量運算資源，故很難導入至行動裝置。為解決此問題，三星 AI 實驗室正研發具有強大效能的微型模型。如此一來，核心研究成果將能運用於所有 AI 技術並實現創新。

問：能否簡單介紹北京三星研發中心和主要工作？

SRC-B 為三星電子 2000 年於中國設立的首間先進研發中心，專注前瞻技術創新，主要研究 AI 及新世代電信技術，包括機器學習、電腦視覺、語言處理、語音智慧及 3GPP 標準化等，並同時致力促成產學合作。AI 實驗室於 2019 年四月成立，負責機器學習的基礎研究與實務應用，並將研究成果導入三星產品。



問：在成功發表主要研究論文並取得其他成就後，目前團隊在執行的計畫為何？

SRC-B 現階段目標是以最有效的方式提升 AI 演算法準確度，同時減少運算複雜性與所需資源量。為此，三星正進行等變網路（Equivariant Network，為幾何深度學習次領域）和動態推論（Dynamic Inference）兩項主題研究，以準確預測並減少所需資料。電腦視覺資料集

包括多樣化的對稱資料，可達到和人眼一樣準確的量測深度，例如影像和光達點雲。於設計等變網路時，三星特別依據資料集本身結構將對稱資料納入，以更少資源實現更優異的表現。

動態推論也是很有趣的研究領域。一般推論方法以固定架構對應所有資料樣本，但動態推論是依據不同資料樣本彈性調整所需的運算資源，簡單樣本使用較少資源，困難樣本則使用較多，以大幅降低平均資源量。

問：從電腦視覺到語音辨識，用戶的應用皆因 AI 基礎研究才得以實踐。能否請您說明為何 AI 基礎研究如此重要嗎？還有您與 AI 實驗室透過哪些研究實現行動體驗最佳化？

網路時代裡資料無所不在，而資料就是知識，欲揭開資料背後的奧祕，並發展出增進大眾福祉的應用，AI 演算法是最有效的工具。

團隊研發出基於資訊瓶頸理論 (Information Bottleneck，假設網路能將含有雜訊的輸入資料去蕪存菁，就像把資料從瓶子擠壓出去一樣，去除不相干的細節) 的網路壓縮演算法，並已應用於影像辨識、圖像分割和機器翻譯等不同任務。我們也積極與 SRC-B 其他實驗室合作研發更強大的 AI 演算法，包括神經網路架構搜索 (Neural Architecture Search，NAS) 和 Once-For-All (OFA) 解決方案。

問：您認為將所有基本行動技術結合運用機器學習的 AI 技術，對用戶來說最主要的效益為何？

基於機器學習的 AI 技術可透過三個面向顯著增進用戶的生活品質：首先，AI 技術是許多便利功能的基礎，例如行動裝置的自動問答系統即為 AI 演算法驅動。其他傳統技術能處理的問題有限，且需預先定義。

相較於僅用傳統技術，加入 AI 技術後能大幅提高許多應用程式的效能。例如相機的神經影像訊號處理 (Image Signal Processing，ISP) 若結合深度神經網路，其影像品質將顯著改善。

最後，AI 技術可提供用戶未曾使用過但相當實用的服務。如 AI 技術可依據用戶的特定偏好開發一套專屬軟體，持續提供升級的裝置體驗。



▲ 中國三星電子研發中心研究團隊

問：團隊如何與中國三星研發中心其他單位互相配合？或甚至是全球其他研發中心？各機構如何整合研究成果，為用戶打造更便利的生活？

我們持續與 SRC-B 其他團隊密切合作，近期我們與視覺運算 (Visual Computing) 團隊共同研究，將基於資訊瓶頸理論的壓縮演算法應用至影像辨識和人像分割，打造尺寸極小但效能絲毫不減的模型。2021 年我們也一同組隊參加電腦視覺與模式辨識會議 (Conference on Computer Vision and Pattern Recognition , CVPR) 的神經網路架構搜索競賽，更透過此解決方案奪得冠軍寶座。

此外，我們也與語言智慧 (Language Intelligence) 團隊合作壓縮機器翻譯模型，成功將其應用程式商品化。

我們堅信，透過與全球 AI 中心的持續交流合作，將創造更進階的研究與應用成果。

問：在您的專業領域裡，目前主要趨勢為何？您如何將熱門技術融入 SRC-B 的研究？

現今熱門的技術很多，如高效網路架構設計、自我監督學習 (Self-Supervised Learning) 及圖形神經網路 (Graph Neural Network) 。

團隊著重於網路壓縮和微型模型設計，十分適合應用於行動裝置。智慧型手機等裝置搭載的運算資源有限，意味著無法容納尺寸過大的行動裝置服務模型，因此我們專注於設計適用於行動裝置的模型。

欲打造輕量化又強大的模型方法有很多種，我們目前主要研究網路剪枝（Pruning）、量化（Quantization）、知識蒸餾（Knowledge Distillation）、神經網路架構搜索和動態推論。

問：您在 SRC-B 這些年來，最自豪的成就是什麼？

與通訊研究（Communication Research）團隊合作下，我們研發出無線通訊專用的 AI 演算法，並於 2021 年中國信通院（CAICT）舉辦的無線通訊 AI 大賽（WAIC），全球 600 多組隊伍中脫穎而出獲得冠軍。能於中國官方 5G+AI 競賽中繳出亮眼成績，我深感驕傲，也更堅信 5G+AI 是深具潛力的研究方向。

【與三星研究院共築未來】
機器學習專家訪問系統軟體專家



嗨 Evgeny，AI 與機器學習的研究成果促成了各類技術的設計與最佳化，技術能順利啟用，系統軟體功不可沒。我想請問，於研發新的軟體系統技術時，您曾遇到哪些挑戰並如何克服？

下一篇文章將帶領讀者與俄羅斯三星研發中心（SRR）系統軟體專家 Evgeny Pavlov 的專訪。