國立中山大學應用數學系
學術演講

講　者：魏澤人 教授 (陽明交通大學智慧計算與科技研究所)

講　題：低成本的 AI 模型微調及應用

時　間：2025/10/16（Thursday）14:10 ~ 15:00

地　點：理 SC 4009-1 教室

茶　會：13:45

Abstract

在生成式 AI 引領浪潮的今天，高昂的運算門檻已成為許多學術單位、中小型企業與個人開發者難以逾越的障礙，訓練甚至是推理時，往往需要龐大的 GPU 叢集。本次演講將探討在消費等級的 GPU 或較為容易取得的單 GPU 工作站/伺服器（如 A100~H200)上，對於生成 AI 的微調訓練及應用。包含現況以及我們相關的研究成果。

我們的一項研究著重於在消費等級 GPU 上強化語言模型的邏輯推理能力訓練。雖然 unsloth 已經提供能在 24GB 記憶體中對 Qwen2.5-3B 模型採用強化學習（GRPO/DeepSeek zero）方式的訓練腳本，但實際使用時，訓練結果並不好。我們探索了不同的訓練策略，最終的訓練配方，能媲美甚至超越過往在 GPU 叢集上的訓練結果。另外一項研究則專注於輕量級視覺語言模型，透過獨特的兩階段微調流程與對資料多樣性的深刻洞察，我們訓練的 3B 模型在圖形介面（GUI）定位任務上的準確率，不僅在同級模型中表現最佳，甚至在部分場景超越了比它大數倍的模型 。

除此之外，在一些產學合作上的經驗，我們也在較低成本及硬體需求的環境下，微調了許多 Video Generative models，客製化達到特定任務，並能在本地端的個人電腦上部署。由於 AI 發展迅速，變化日新月異，但至少在現階段，我們將探討在資源限制下，AI 模型訓練及開發仍有的許多可行性及潛力。

敬　請　公　告！　歡　迎　參　加！

應用數學系：http://math.nsysu.edu.tw

校園地圖：http://math.nsysu.edu.tw/var/file/183/1183/img/779/nsysu_math_map.jpg

交通資訊：https://www.nsysu.edu.tw/p/412-1000-4132.php?Lang=zh-tw

應用數學系　　校園地圖　　交通資訊