

小模型，大貢獻：準確且高效的中文新聞摘要模型訓練研究

Small Model, Big Impact: A Study on Training for Accurate and Efficient Chinese News Summarization Models

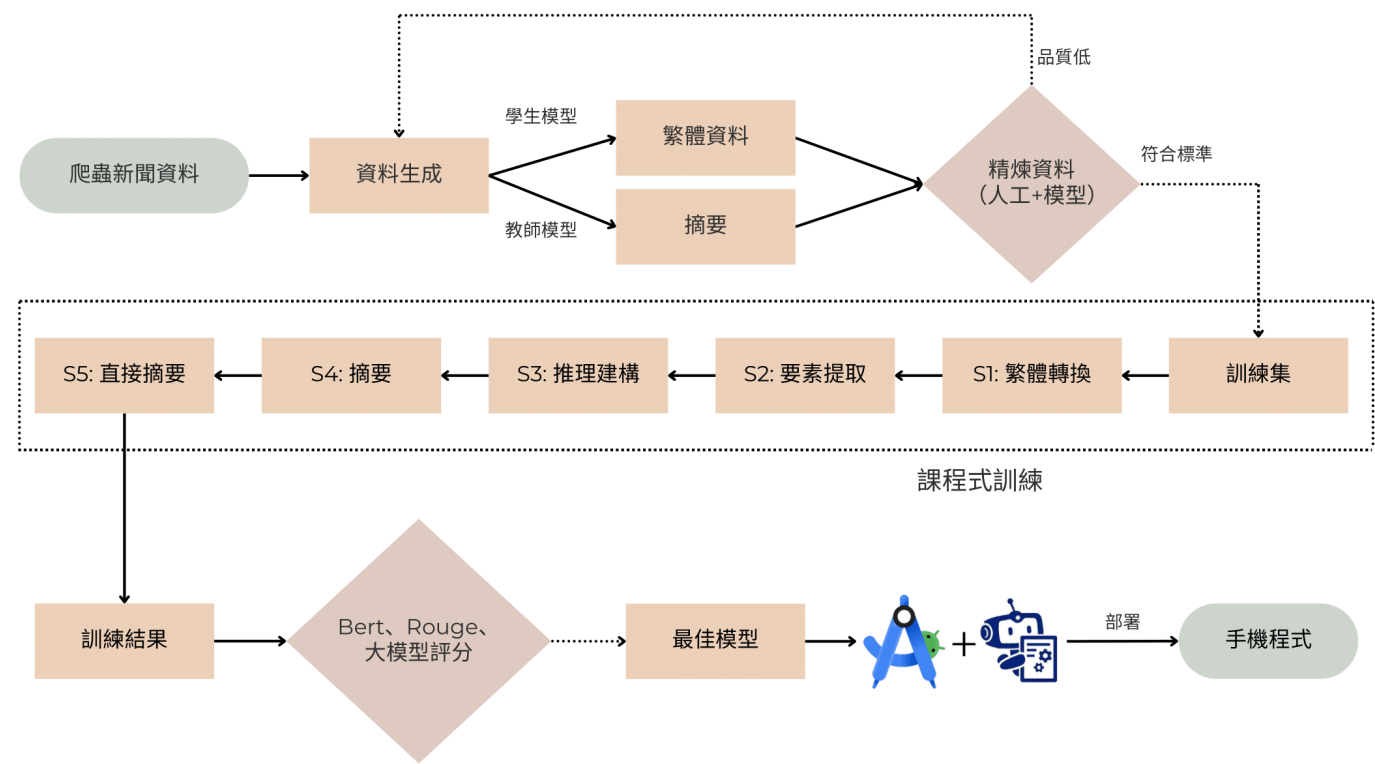
指導教授：陳響亮
專題成員：王郁豪
開發工具：Android Studio、Python、mlc-llm
測試環境：Android 9.0 (arm64-v8a)、
Ubuntu 24.04.1 LTS (x86_64)

一、簡介：

在資訊爆炸的現代社會中，讀者常因時間有限與聾動標題影響，難以正確理解事件核心。為此，本研究旨在尋找最佳的訓練策略，並開發一個能在手機等資源受限設備上運行的中文新聞摘要模型，協助用戶快速獲取重點資訊。

我採用 Qwen2.5-0.5B-Instruct 作為學生模型，並以 Qwen2.5-32B-Instruct 作為教師模型，透過模型蒸餾與課程式訓練策略，逐步引導學生模型學習繁體中文轉換、要素提取、推理建構與新聞摘要等能力。訓練過程分為五個階段，強調從簡至難的學習過程，同時針對資料生成方式、訓練策略與模型參數進行調整與實驗，以持續提升模型泛化與生成品質。

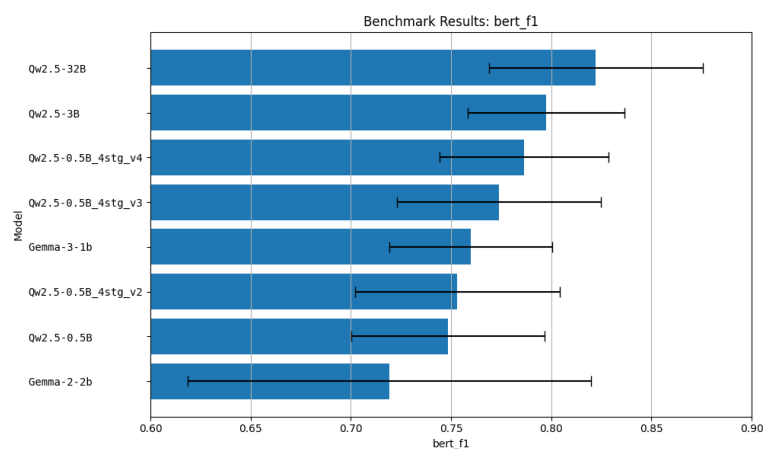
最終成果顯示，學生模型不僅能有效生成內容正確、語句流暢的繁體中文摘要，亦能修正教師模型在繁體中文用語上的部分錯誤，展現良好的應用潛力。



圖一：實驗流程圖

二、測試結果：

1. 資料生成策略方面，實驗發現教師模型不擅長以推理方式直接生成摘要，易產生不連貫或失焦內容。相較之下，先生成摘要再反推出要素與關聯表現較佳。
2. 在訓練策略方面，採用線性遞減學習率（learning rate decay）訓練效果效果較差；此外，與 LoRA 或凍結部分參數的微調策略相比，直接微調全參數模型的表現優於其他方法。
3. 在大規模訓練資料下，課程式學習對生成品質提升有限，但能提升繁體中文生成的準確率。而在僅使用約 10 萬篇新聞預訓練、並以少量資料微調的自訂義小模型中，課程式學習顯著提升模型的摘要準確度。
4. 整體而言，隨著訓練策略與資料生成流程的持續優化，學生模型生成的摘要更加精煉外，也避免出現教師模型的問題(如異常 token、錯誤譯字等)。整體生成品質接近大型模型，但參數量更小、執行速度更快，實現輕量化與高效能並存的目標。



圖二：不同學生模型、教師以及相近參數的模型之 BERTScore

最新新聞

法網／澤瑞夫男單首輪晉級 梅德維夫鏖戰5盤惜敗
德國名將澤瑞夫 (Alexander Zverev) 今天在法網男單首輪以直落三盤敗對手，順利晉級次輪。...

大樂透頭獎連6槓 下期獎金上看1.9億元
大樂透第114056期今晚開獎。中獎號碼為32、29、26、17、40、16，特別號10；派彩結果，...

AI升級！三星智慧電視超大尺寸 冰箱有螢幕幫你記食材
AI技術的快速進化，全面影響了現代人的日常生活。三星電子今日在台發表多款2025年全新智慧家電，涵蓋...

涉性騷教授請辭婦產科醫學會理事長 會員代表:影響不大
台大婦產部不孕症權威醫師C姓教授遭性騷擾住院醫師，經院方調查判定性騷擾成立且情節嚴重，後續遭台大處...

中職／王威晨延長賽敲超前二壘打 兄弟擊敗雄鷹6連勝
中職中使兄弟今天與台鋼龍鷹交手9局打完僵持不下，兄弟隊王威晨10局上敲出超前比數二壘打，率隊以4比3...

印度4月銷美iPhone大增逾七成 大陸竟反降
根據市場研究公司數據，4月由印度銷往美國的iPhone年增76%；同期大陸製造的iPhone對美銷量...

川普砍哈佛經費再出招 擬取消30億元聯邦機構合約
總統川普政府兩位高階官員告訴有線電視新聞網 (CNN)，聯邦政府將指示所屬機構取消與哈佛大學 (Harv...

prefill: 2.3 tok/s, decode: 5.0 tok/s

投完在4：4平手局面退場，吳承諭第5局接手面對6打席，被敲3安打、1保送，悍將靠林澤彬、王正棠的安打各添1分，取得6：4領先；悍將5局打完已經敲出12支安打，且完成先發九棒全員安打。悍將由黃保羅先發6局正好用100球，被敲6安打，失4分都是因全壘打，分別是第1局林安可三分砲、第5局林靖凱陽春砲，也恰好都是兩人的生涯500安。悍將第7局林拱呈登板，被陳重羽敲二壘打、保送林靖凱後退場，李吳永勤強勢連抓3個出局數「拆彈」成功，守住當時僅差距2分的緊繃戰局；7局下悍將再度攻占滿壘，王正棠打向二壘方向的滾地球，全力衝上一壘避免雙殺延續攻勢，申皓瑋後續的安打再擴大領先優勢。

Show as Markdown

富邦悍馬今天以無雙模式打敗統一，擊敗統一獅隊，最終以8：4勝，終結4連敗。豪馬在本季打獅隊前7戰全敗後，拿下對戰第1勝，今年只剩面對樂天桃猿隊 (0勝8敗) 還沒拿到勝場。十勝之下，蒙德茲投手生涯成績比肩高國麟，雙打比肩。

圖三：手機程式畫面：使用者可選取文章 → 手機端生成摘要 → 顯示摘要結果