

Machine Learning Assignment 10-313652008

313652008 黃睿帆

November 8, 2025

Problem 2: AI 的未來與機器學習的基石

我認為目前 AI 無法做到，但 20 年後有可能做到的事情是 **AI 理解與重構人類記憶**。(P.S. 此作業有部分經由 Chat GPT 潤飾語句和分析需用到那些機器學習之原理)

1. AI 二十年後的未來能力：理解與重構「人類記憶」

在 20 年後，AI 可能具備「從人腦活動中理解並重構人類記憶內容」的能力。這項能力指的是：AI 能夠透過腦部掃描資料（例如 fMRI、EEG、神經電極訊號），學習辨識大腦在回想畫面、聲音或情緒時的神經活動模式，並將這些模式轉換成具體的可視化影像、聲音或文字敘述。

我的動機和靈感來源其實是多年以前看過的一個劇集「神探伽利略」，第二季內有某個案子，主角物理學家湯川教授提到了有個高維度矩陣研究中心，這個機構能將人類腦部的影像成功映像化來抓犯人，雖然後面揭密這是假的，那是教授胡謔的，將人腦內影像映像化以當時技術來說根本還辦不到（註：1990 年代電腦才剛能做出蒙太奇效果，但也感嘆科技的進步，三十五年後電腦已經具有強大的 AI 功能），他是故意說的來擾亂真正的犯人來讓犯人露出馬腳。

因此這給了我一個啟發，或許未來的 AI 能夠構建出人腦中的影像，而且還能夠重構人類的記憶，用於醫學上治療失憶或有創傷後症候群症狀的患者，也能像上面的劇集一樣，讓目擊證人重構案發現場的畫面，來抓犯人並釐清真相。

更具體來說，從人腦活動中理解並重構人類記憶內容的應用場景包括：

- 醫療復健：幫助失智症患者或創傷後壓力症候群（PTSD）患者「喚回」或「重整」破碎的記憶，協助心理治療與情緒修復。
- 司法鑑識：在倫理審核與個人同意下，協助還原目擊者或被害人的事件記憶，提高證詞準確性。
- 人機介面：讓人類能直接以「想像」與 AI 溝通，例如透過想像畫面來控制電腦、生成藝術或虛擬世界。
- 文化與歷史保存：在未來，人類可能能「記錄與分享」自身記憶，形成全新的個人與文化檔案形式。

為什麼重要：因為這項映像化技術能讓我們真正理解「人類思考與記憶的運作方式」，突破語言與表達的限制。它將徹底改變醫療、教育、藝術與人文研究的模式，但也同時也挑戰了「意識」與「個人隱私」的界線。因此使用時必須小心以及更加謹慎...

2. 此技術所牽涉的機器學習類型與可能理由

這項技術所使用的主要機器學習方法可能包含：

- 監督式學習 (Supervised Learning)
- 生成式模型 (Generative Models) (如 Diffusion Models、Variational Autoencoders、GANs)
- 多模態學習 (Multimodal Learning)
- 表徵學習 (Representation Learning)

多模態學習是一種深度學習方法。它整合和處理多種類型的資料，例如文字、音訊、圖像或影片。這些不同類型的資料叫做模態。這種整合能夠更全面地理解複雜資料，從而提高模型在視覺問答、跨模態檢索、文字到圖像生成、美學排名和圖像字幕等任務中的效能。(資料來源: 維基百科)

特徵學習 (feature learning) 或表徵學習 (representation learning) 是學習一個特徵的技術的集合：將原始資料轉換成能夠被機器學習來有效開發的一種形式。它避免了手動擷取特徵的麻煩，允許電腦學習使用特徵的同時，也學習如何擷取特徵：學習如何學習。(資料來源: 維基百科)

(註: 這兩項似乎在課堂上沒有琢磨太多，但這兩項是 Chat GPT 認為此項技術會用到的機器學習技巧)

3. 為什麼需要這些學習類型：

- 監督式學習：要讓 AI 學會「神經活動模式」和「視覺或語意內容」的對應關係，必須以已知配對資料訓練，像是 (1) 輸入資料 (人在觀看或想像特定影像時的大腦神經訊號)；(2) 標籤資料：對應的實際影像內容。透過監督學習，AI 能逐漸掌握不同腦區活動與記憶內容之間的對應特徵。
- 生成式模型：因為人腦的記憶重構不是單純分類，而是「從腦部訊號生成視覺或語音內容」。所以生成模型 (如 diffusion model 或 VAE) 或許能將潛在的腦部特徵向量轉換成影像或聲音，重建人類曾見或想像過的場景。
- 多模態學習：記憶往往包含多種感官成分 (影像、聲音、情緒)。所以 AI 必須學會在不同感官模態間對齊與整合資訊，建立統一的「記憶表徵空間」。
- 表徵學習：同時，為了讓 AI 能理解「記憶」這種高維、抽象概念，它需透過表徵學習在神經信號與語意之間建立橋樑，學習穩定的潛在特徵 (latent representations)。

4. 可能的「資料來源」與「目標訊號」

- 資料來源 (Input Data): 受試者在回想、觀看或夢境狀態下的神經影像數據 (fMRI、EEG、ECoG 等)。
- 目標訊號 (Output Target): 對應的記憶內容：影像 (視覺場景)、語音 (對話片段)、情緒標籤或文本描述。
- 訓練資料型態: 神經訊號與感知內容的配對資料集。
- 學習回饋與互動: 模型可透過「重建誤差」作為回饋信號 (例如生成影像與原始影像的相似度)，屬於靜態監督學習；若加入使用者互動 (例如主觀評價記憶重現的準確性)，則屬於半監督或互動式學習。

5. 第一步的「模型化」：簡化模型問題 (model problem)

任務摘要: 以 EEG (或低場強度、常規 fMRI 若有條件) 記錄受試者觀看 500 張 128×128 自然影像 (每張重複觀看若干次)，訓練一個模型把 EEG 時序特徵映射到一個簡單的影像潛在空間 (自訓練的小型自編碼器 latent)，最後用生成器重建低解析度影像 (64×64) 並評估重建品質與識別率。簡而言之，把腦訊號 → 潛在向量 → 生成影像，目標是重建真實觀看的影像。(註: 參考 Chat-GPT)

為什麼這個簡化問題能代表最終能力？因為最終希望 AI 能理解並重構影像或是記憶，核心是「從神經活動恢復高層語意/感官或是影像內容」。因此專注於這一核心：從神經活動推回能描述圖像語意與細節的 latent 表徵。並採用預訓練生成器 (G) 把生成責任交給成熟的影像模型，能把挑戰聚焦在「神經 → 語意潛在表徵」的映射，這正是未來記憶重構系統必須解決的關鍵子問題。(以下是擴展:) 若此映射可行，可延伸到更高解析度、更多模態 (聲音、情緒)、或跨受試者通用化，逐步接近完整影像和記憶重建。

可能需要的數學工具:

- 監督式迴歸 / 映射：將神經特徵回歸到 latent vector。可用的模型：線性回歸、Ridge、MLP、Transformer、或條件神經網路。
- 表徵學習：使用 CCA、Deep CCA、多視角對比學習 (contrastive learning) 來對齊 neural 與 image latent 空間。
- 生成式模型作為解碼器：預訓練 VAE / GAN / Latent Diffusion，從 latent 產生影像。

- 損失設計：組合 pixel/LPIPS (perceptual loss)、latent L2 loss (若真實 latent 可得)、對抗損失 (提升真實感)、以及 CLIP 相似度損失 (強化語意一致性)。
- 不確定性估計：貝葉斯回歸或 MC-dropout 來量化 latent 推斷不確定性 (important for 安全與倫理)。
- 統計推斷和檢定：bootstrap、permutation test、FDR control (多重比較修正)。
- 逆問題與正則化：分析性地把模型視為 illposed inverse problem，使用 Tikhonov regularization、sparsity priors 或 manifold constraints 來穩定解。
- 資料擴增 / 合成驗證：若資料有限，可用合成神經訊號模擬 (simulate neural responses from known image features) 做可行性檢測。

小結

AI 理解與重構人類記憶的能力，結合了深度學習、神經科學與生成式模型的進展。它的實現不僅需要巨量神經影像資料與多模態對應訓練，也需要發展新的倫理規範與隱私保護機制。這不只是技術挑戰，更是人類對「自我意識」與「記憶」本質的重新探索。