

壹、方法設計

本專案旨在預測社群平台中貼文的觀看數，資料來源為平台真實貼文資料，包含圖像、標題文字、發文時間與分類等多種特徵。此任務屬於典型的多模態回歸問題，挑戰在於有效整合結構化資料與非結構化資料（如圖像與文字）。

經由同學的介紹我使用了 AutoGluon 套件中的 MultiModalPredictor (AutoMM) 作為主力模型。AutoMM 能夠同時處理圖像、文字與表格欄位，並自動選擇最佳 backbone（例如 ViT、BERT、小型卷積網路等）進行訓練與調參，適合本任務的資料特性與目標。

預測流程重點如下：

(一)Label 平滑處理：使用 $\text{label_log} = \log(1 + \text{label})$ 處理目標變數，有效減緩長尾分布的影響。

(二)結構化特徵設計：在原始特徵基礎上，額外加入：

title_length：標題長度

is_weekend：是否為假日

hour_bin：時段分群（凌晨、上午、下午、晚上）

(三)模型訓練：使用 AutoMM，設定 $\text{time_limit} = 1800$ ，預設使用 high_quality 模式，確保模型表現與穩定性。

(四)預測與反轉：預測後使用 $\text{expml}()$ 將結果還原為原始觀看數。

貳、引用方法與差異說明

本次實作主要參考 AutoGluon 官方範例教學 (AutoMM 多模態分類與回歸模型) ，

但依據本任務資料屬性與目標進行如下修改：

項目	原教學	本作業做法
預測任務類型	分類任務	回歸任務 (觀看數)
Label 處理	原始值	使用 <code>log1p(label)</code> 平滑
結構化欄位	僅使用內建欄位	額外設計 <code>title_length</code> , <code>hour_bin</code> 等
模型輸出方式	直接預測	預測 <code>log</code> 值後使用 <code>expm1()</code> 還原
模型類型	預設 backbone	由 AutoMM 自動挑選並微調

參、實驗結果與分析

AutoMM 訓練期間觀察模型在驗證集上的 RMSE 表現，顯示模型學習穩定且有效：

Epoch	val_rmse (最佳)
0	0.89890
1	0.87576
2	0.82706
3	0.78212
4	0.78167
5	0.78123
6	0.75964

此外，將此模型輸出上傳至平台 leaderboard，public Scoreu 顯著優於 baseline，且表現穩定。這表示 AutoMM 成功學習並整合了圖像、文字與時間結構等不同模態的訊號。

我們也觀察到，log(label) 處理大幅降低了極端值對訓練過程的影響，能夠有效改善過擬合；而結構欄位如 title_length, hour_bin 對模型收斂速度與準確性也有明顯助益。

肆、參考資料與引用來源

AutoMM 教學文檔

<https://auto.gluon.ai/stable/tutorials/multimodal/index.html>

AutoMM 論文

Zeng et al., AutoGluon Multimodal, NeurIPS 2022

結構特徵設計靈感：參考常見社群資料建模實務與時間特徵工程

伍、後續優化方向建議

嘗試 pseudo-labeling：將高信心 test 結果納入訓練迴圈

加入特徵重要性分析（如 SHAP）理解模型關注重點

嘗試更輕量化 backbone（MobileViT、TinyBERT）提升訓練效率

嘗試進一步 tuning backbone 結構或層級凍結策略