

Term Project

2024 NTHU Natural Language Processing

Hung-Yu Kao

IKM Lab TAs

Term Project 規則

- 本次 Term project 提供五個 Kaggle 競賽任務
- 請各組從五個任務中自行選擇其中一項來進行 Term project
- 你們需要：
 - 統整兩個以上的 code (這五個 Kaggle 任務都已經有許多存在的code)，並產生自己的作法
 - 分析自己作法和已經存在的code作法相比的優勢與劣勢
 - 提出新的做法
 - 努力取得前面的排名，你們組在Leaderboard上的排名將納入Term Project評分，本次競賽皆為Code competition，因此需要仔細閱讀競賽規則，建立好程式碼後，上傳到 Kaggle

Outline of tasks

*Baseline code/notebooks can be found.

平台	編號	競賽名稱	資料類型	任務類型	每日上傳次數
Kaggle	1	LLM Classification Finetuning (link)	文字	分類	10
	2	Learning Agency Lab - Automated Essay Scoring 2.0 (link)	文字	分類	5
	3	LLM Prompt Recovery (link)	文字	檢索	5
	4	LLM - Detect AI Generated Text (link)	文字	分類	5
	5	Eedi - Mining Misconceptions in Mathematics (link)	序列、數值	迴歸	5

1. LLM Classification Finetuning

- [頁面連結](#)
- 任務簡介
 - 輸入: prompt 以及 a 模型、b 模型根據 prompt 生成的回應
 - 輸出: a 模型回應較好的機率、b 模型回應較好的機率以及兩模型回應一樣好的機率
 - 三個機率值總和為 1
- 評估方式:
 - 三個預測機率值與正確答案的 **Log Loss**
- 每日上傳次數：10 (競賽已結束，但仍可以上傳)

2. Learning Agency Lab - Automated Essay Scoring 2.0

- [頁面連結](#)
- 任務簡介
 - 輸入: 文章
 - 輸出: 預測該文章的分數 (1 - 6)，分數皆為整數
- 評估方式:
 - 預測與正確答案的 **Quadratic Weighted Kappa** (具體算法請見頁面連結)
- 每日上傳次數：5 (競賽已結束，但仍可以上傳)

3. LLM Prompt Recovery

- [頁面連結](#)
- 任務簡介
 - 輸入: 文章段落，利用LLM進行風格轉換後的文章段落。
 - 輸出: 預測風格轉化任務時提供給LLM的prompt (Exp: Rewrite this essay with style of J.K. Rowling)。
- 評估方式:
 - 評估方式為predicted prompt和actual prompt之間的**sharpened cosine similarity**分數。
- 每日上傳次數：5 (競賽已結束，但仍可以上傳)

4. LLM - Detect AI Generated Text

- 頁面連結
- 任務簡介
 - 輸入: **文章的内容** (包含了學生或 LLM 的寫作内容) , 以及每篇文章對應的題目背景
 - 輸出: 判斷這篇文章**是否由 LLM 生成** (0 表示由學生撰寫 , 1 表示由 LLM 生成)
- 評估方式:
 - 以 **ROC 曲線下面積 (AUC , Area Under the Curve)** , 評估對於正確答案的預測效果
- 每日上傳次數 : 5 (競賽已結束 , 但仍可以上傳)

5. Eedi - Mining Misconceptions in Mathematics

- [頁面連結](#)
- 任務簡介
 - 輸入: 多選題題目敘述、四個選項
 - 輸出: 各個選項對應到的錯誤概念 (對應表共2586個類別) , 可預測最多25個
- 評估方式:
 - 預測與正確答案的 **Mean Average Precision @ 25 (MAP@25)** (具體算法請見頁面連結)
- 每日上傳次數 : 5
- 競賽時間:
 - 2024/9/12 - 2024/12/12

Term Project Rating

項目時間點 (check point)	Date	繳交資料	Grade
CP0: Task selection	11/17	Task preference form	
CP1: Presentation (offline)	11/30 (W13)	報告簡報檔案 (pdf), Youtube clip (10-15 mins)	7%
CP2: Review	12/8 (W14)	Review + Scoring	3%
CP3: Oral Presentation (30 teams) /Poster	12/10, 12/17 (W15, W16)	報告(10mins)/poster 簡報檔案 (pdf)	5%
CP4: Final Report + score ranking	12/27 (W17)	文書報告檔案 (.docx), code pack	12% + 3%