2024 Information Retrieval and Extraction HW3

Dialogue-Based Photo Retrieval

Task Introduction

With the rapid rise in popularity of instant messaging tools (such as LINE and Messenger) in recent years, sharing photos has become an indispensable part of online conversations. In this assignment, you need to select the most relevant image based on the textual content of each online dialogue.

- Measure photo relevance to a dialogue context
  - Use any method to encode images or both images and text into the same space to compare the similarity between images and dialogue content.
- Requirement
  - Upload your submission to Kaggle
  - Submit a report and your source code to E3

Dataset

- train.jsonl
    - Contains dialogues and corresponding image IDs needed for training
    - link
- train_images
    - Contains train images needed for training
- test.jsonl
    - Contains dialogues that need to be used for prediction
    - link
- test_images
    - Contains test images needed for prediction
- test_images.jsonl
    - Contains test image informations and corresponding image IDs
    - link

Training Data

train.jsonl - each line is a json dict and contains following attributes:

- **dialogue** - List[Dict] the content of dialogue
  - **share_photo** - Boolean value denoting whether a photo is shared in this turn.
  - **user_id** - 0 or 1. User ID of this turn.
  - **message** - Text of one conversation turn. Empty when share_photo is true.
- **dialogue_id** - Integer. Unique dialogue id.
- **photo_description** - String. Photo description. It includes info about object labels in the photo.
- **photo_id** - Image ID of the photo.

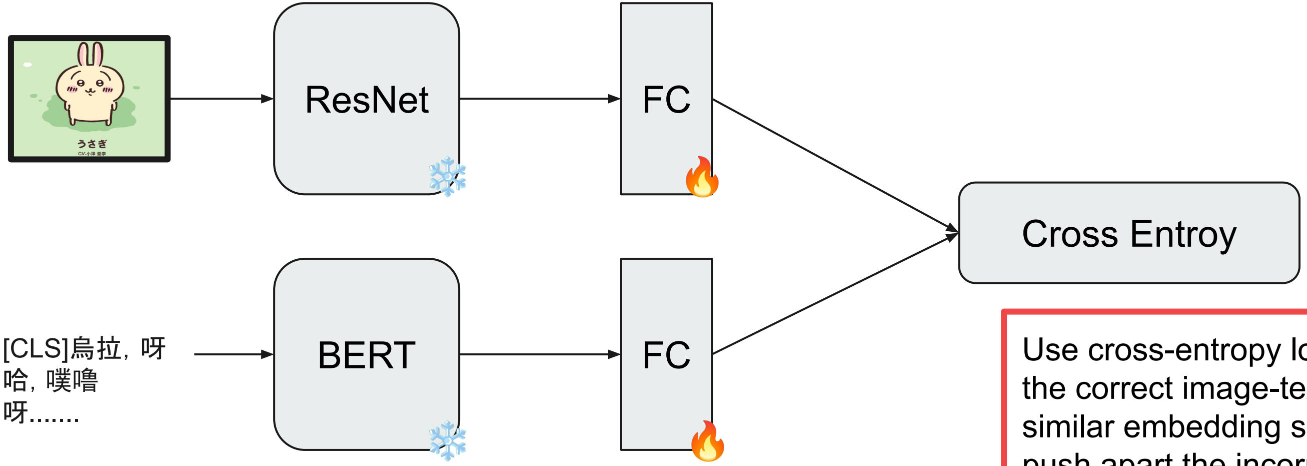
Testing Data

test.jsonl - each line is a json dict and contains following attributes:

- **dialogue** - List[Dict] the content of dialogue
  - **share_photo** - Boolean value denoting whether a photo is shared in this turn.
  - **user_id** - 0 or 1. User ID of this turn.
  - **message** - Text of one conversation turn. Empty when share_photo is true.
- **dialogue_id** - Integer. Unique dialogue id.
- Please note that when making predictions for the Test set, you only need to consider possible recommended images from **test_image**

Method 1 - Dual Encoder

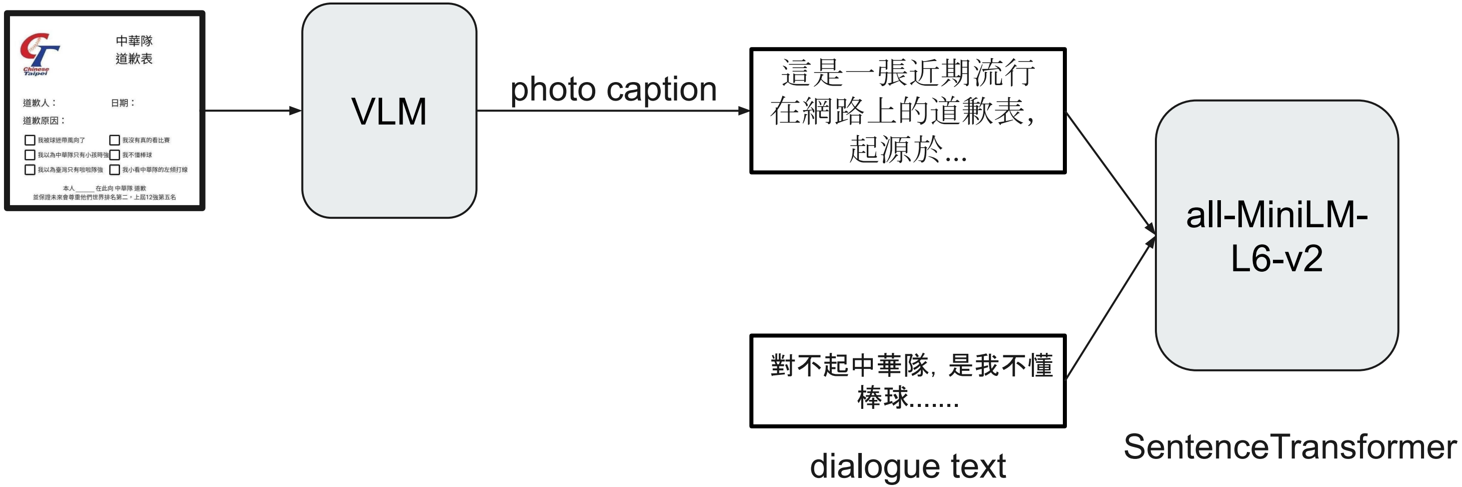
- Use an **pretrained** image encoder and a text encoder to encode data from both modalities, image and text, into the same embedding space



Use cross-entropy loss to map the correct image-text pairs to a similar embedding space and push apart the incorrect pairs.

Method 2 - VLM captioning

- Use a Vision Language Model(VLM) to generate a caption for each photo and compare the similarity between the caption and the dialogue text.



Chinese Taipei

中華隊
道歉表

道歉人：　　日期：

道歉原因：

□ 我被球迷帶風向了
□ 我以為中華隊只有小孩時強
□ 我以為臺灣只有啦啦隊強
□ 我沒有真的看比賽
□ 我不懂棒球
□ 我小看中華隊的左傾打線

本人______在此向 中華隊 道歉
並保證未來會尊重他們世界排名第二，上屆12強第五名

VLM

photo caption

這是一張近期流行在網路上的道歉表,起源於...

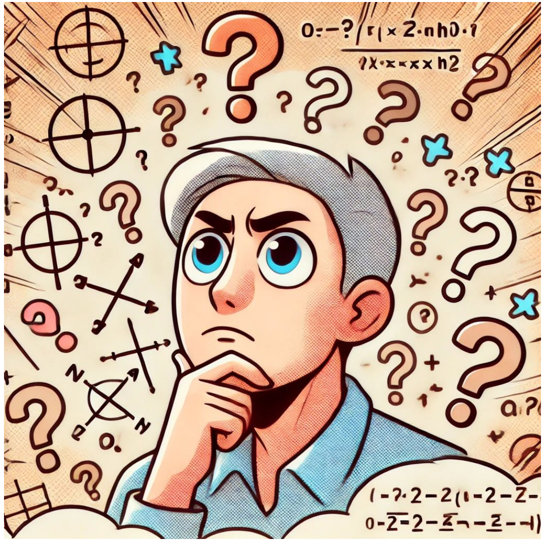
對不起中華隊, 是我不懂棒球.......

dialogue text

all-MiniLM-L6-v2

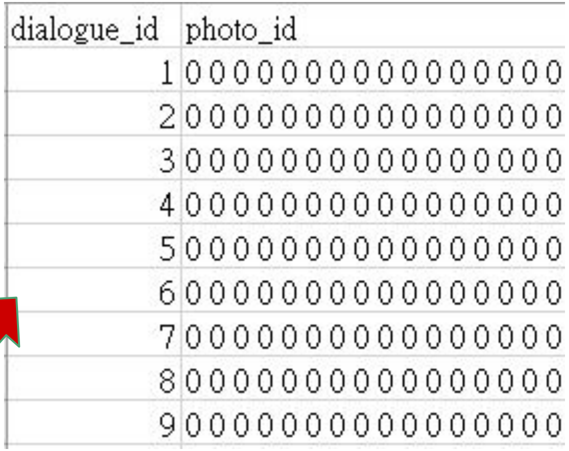
SentenceTransformer

Method 3 - Any reasonable way you can think



Kaggle

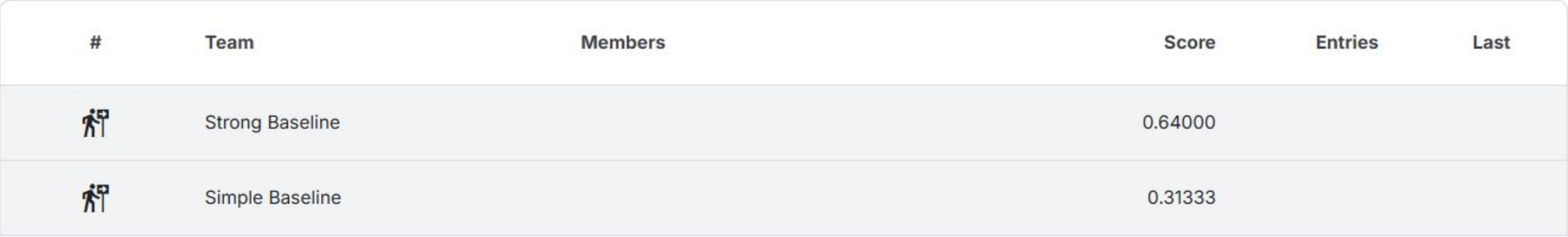
- Kaggle link
- Display team name : <student ID>
- Submission format
  - A 601*2 .csv file, first row is for the column name and the last 600 rows for your result.
  - Column name must be **dialogue_id** and **photo_id**.



dialogue_id	photo_id
1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
3	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
4	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
5	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
6	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
7	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
8	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
9	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

30 Image IDs separated by **spaces**

- There is one simple baseline and one strong baseline. Beat them to achieve a higher score.



#	Team	Members	Score	Entries	Last
	Strong Baseline		0.64000		
	Simple Baseline		0.31333		

Kaggle(cont.)

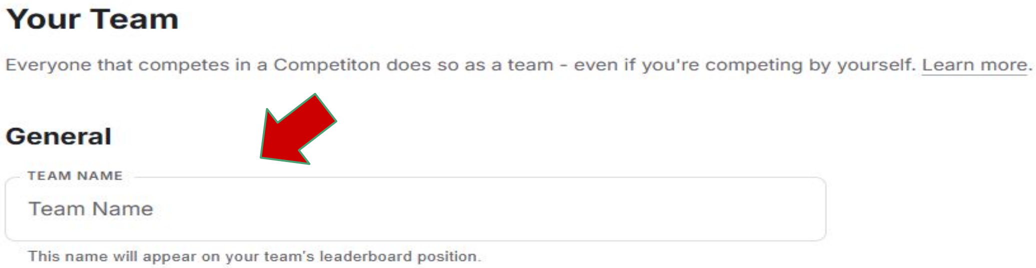
- The scoring metric is Recall@30.
- You can submit at most 5 times each day.
- You can choose 3 of the submissions to be considered for the private leaderboard, or will otherwise default to the best public scoring submissions.
  You can only view your private leaderboard score after the competition has ended.
- Public leaderboard is calculated with 50% of the test data, and private leaderboard is calculated with other 50% of the test data, so the final standings may be different.
- Please tune your model parameters using your own validation set instead of adjusting parameters based on the public leaderboard. Otherwise, it's easy to overfit, leading to poor performance on the private leaderboard.

Change your team name

2024 Information Retrieval & Extraction Homework3

Measure image relevance to a dialogue

Remember to change the team name to <student ID>, or there will be a deduction of 5 points for HW 3.



Your Team

Everyone that competes in a Competiton does so as a team - even if you're competing by yourself. Learn more.

General

TEAM NAME

Team Name

This name will appear on your team's leaderboard position.

Report Submission

Answer the following 3 questions:

1. What kind of pre-processing did you apply to the photo or dialogue text? Additionally, please discuss how different preprocessing methods affected the performance of the models?
2. How did you align the photo and dialogue text in the same embedding space? Use pretrained model or train your own?
3. Please discuss based on your experimental results. How do you improve the performance of your model? (e.g. add a module or try different models and observing performance changes). What was the result?

Please answer the questions in detail to receive full points for each question.

Grading policy

- Kaggle (70%)
  - 30% based on the public leaderboard score and 70% based on the private leaderboard score
  - Leaderboard score consists of basic score and ranking score
    - Basic score :

      Over strong baseline : 55

      Over simple bassline : 40

      Under simple baseline : 25
    - Ranking score:

      15-(15/N)*(ranking-1), N=numbers of people in the interval
- Report (30%)
  - 10 for each quesiton

E3 Submission

Submit your source code and report to E3 before 12/20 (Fri.) 23:59.

No late submission !

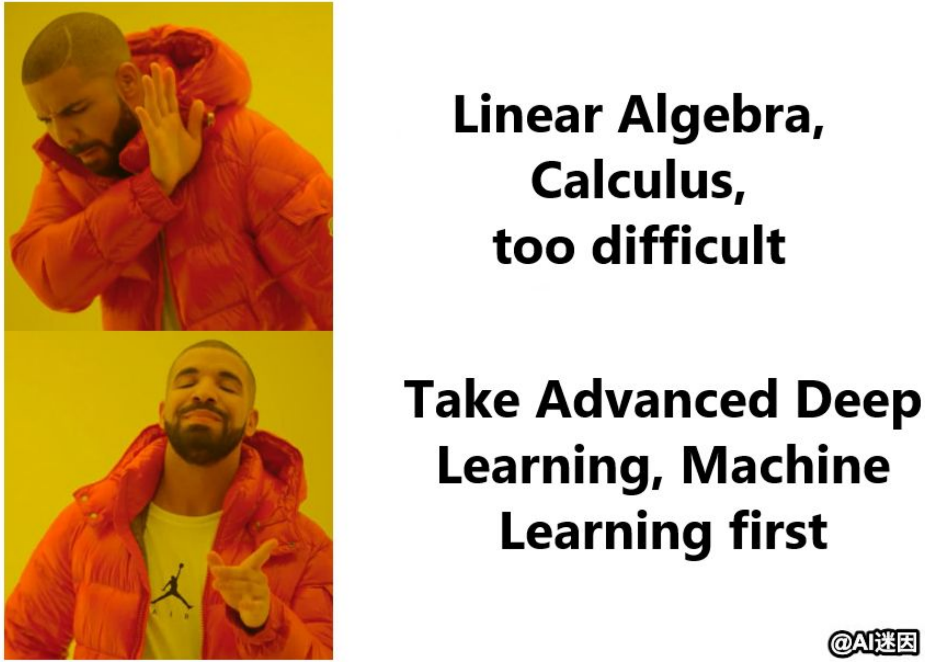
Follow the submission format or there will be a deduction of 5 points for HW 3 !

- Format
  - source code : HW3_<student ID>.py or HW3_<student ID>.ipynb
  - report : HW3_<student ID>.pdf

If you have any question about HW 3, please feel free to contact with TA : CHENG-XIN SONG

through email chengxin0913.cs12@nycu.edu.tw

Have Fun !


Linear Algebra,
Calculus,
too difficult
Take Advanced Deep
Learning, Machine
Learning first
@AI迷因