Eedi - Mining Misconceptions in Mathematics

Group 4

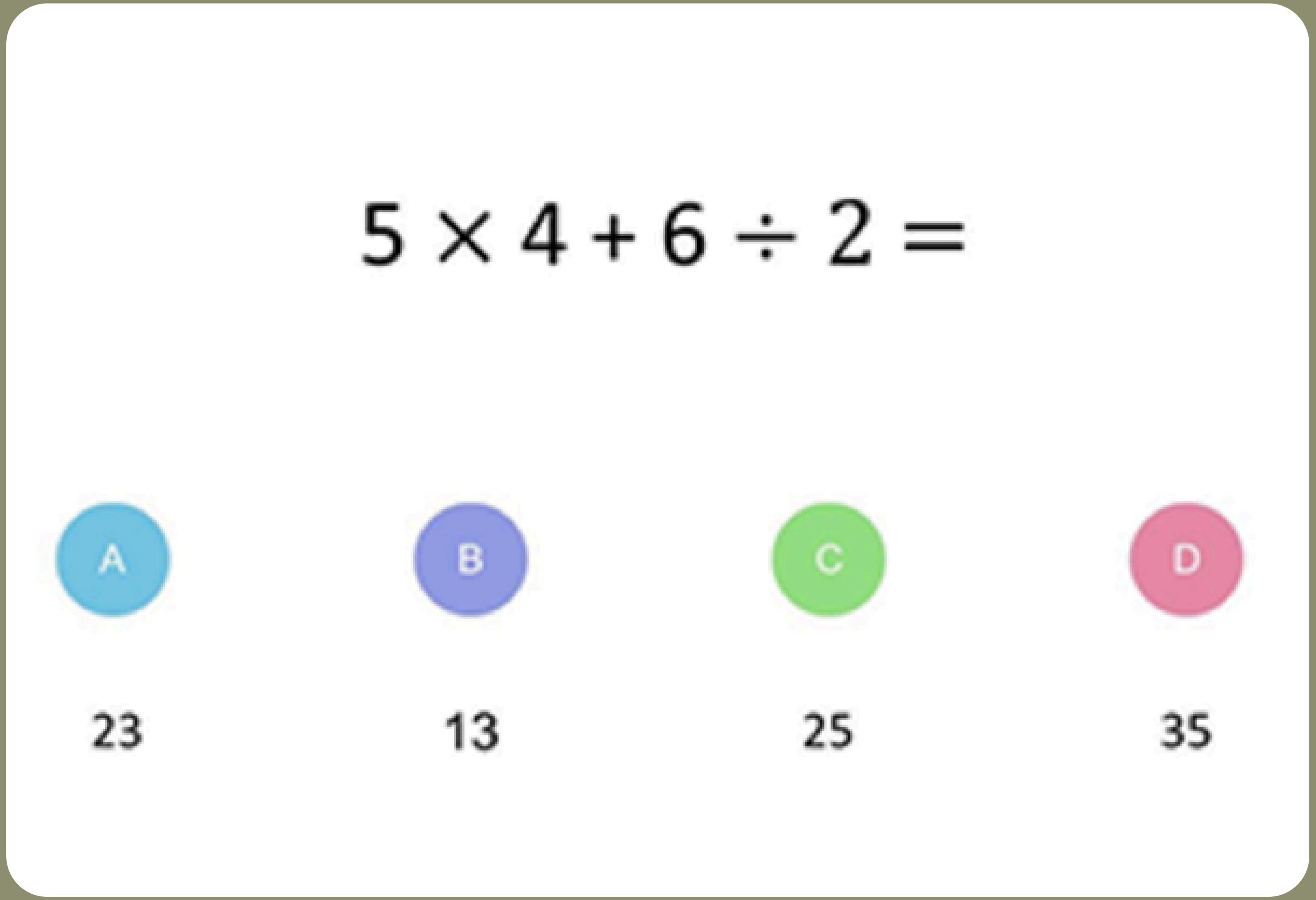
NCKU 許宸華
NCKU 蕭力文

Overview

You'll develop an NLP model driven by ML to accurately **predict the affinity between misconceptions and incorrect answers** (distractors) in multiple-choice questions.

If a student selects the distractor "13,"

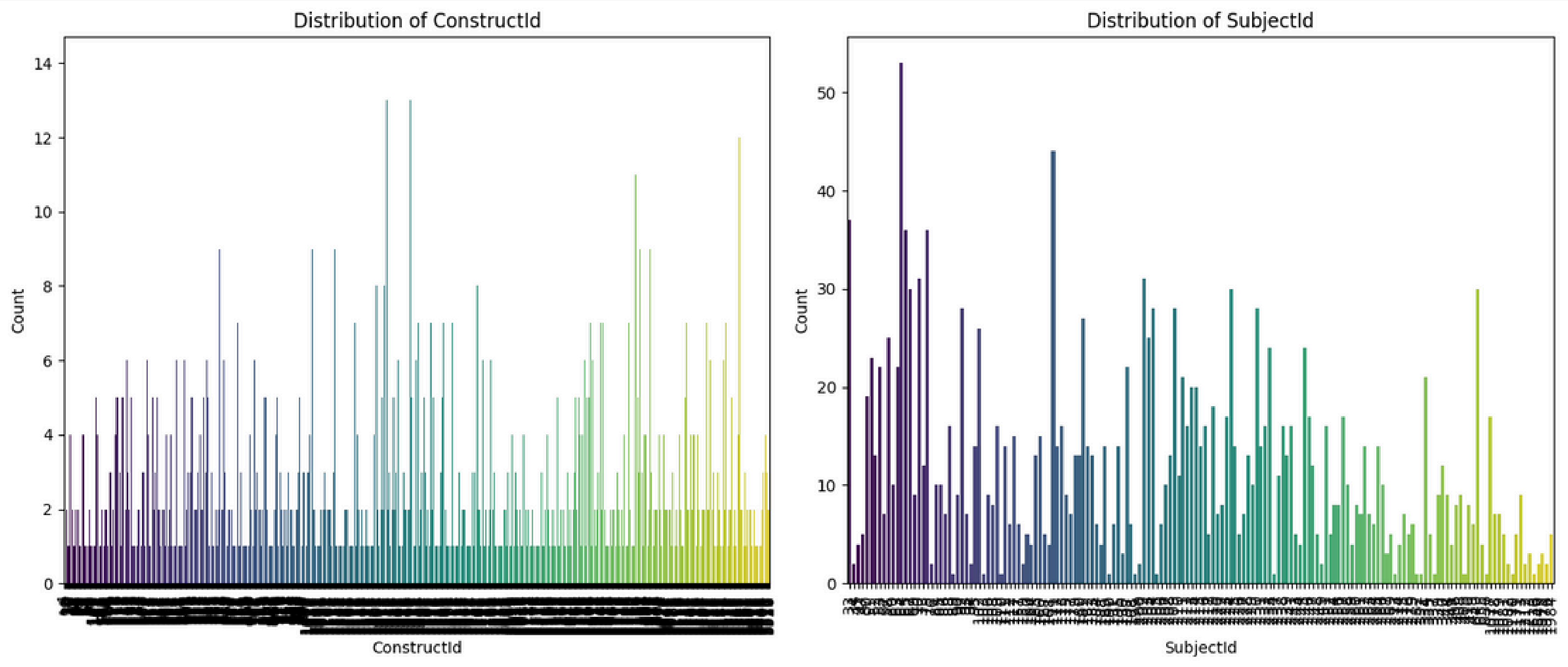
Misconceptions：
"Carries out operations from left to right regardless of priority order."



About Dataset

[train/test].csv:

- QuestionId - Unique question identifier (int).
- ConstructId - Unique construct identifier (int) .
- ConstructName - Most granular level of knowledge related to question (str).
- CorrectAnswer - A, B, C or D (char).
- SubjectId - Unique subject identifier (int).
- SubjectName - More general context than the construct (str).
- QuestionText - Question text extracted from the question image using human-in-the-loop OCR (str) .
- Answer[A/B/C/D]Text - Answer option A text extracted from the question image using human-in-the-loop OCR (str).
- **Misconception[A/B/C/D]Id** - Unique misconception identifier (int). Ground truth labels in train.csv; your task is to predict these labels for test.csv.



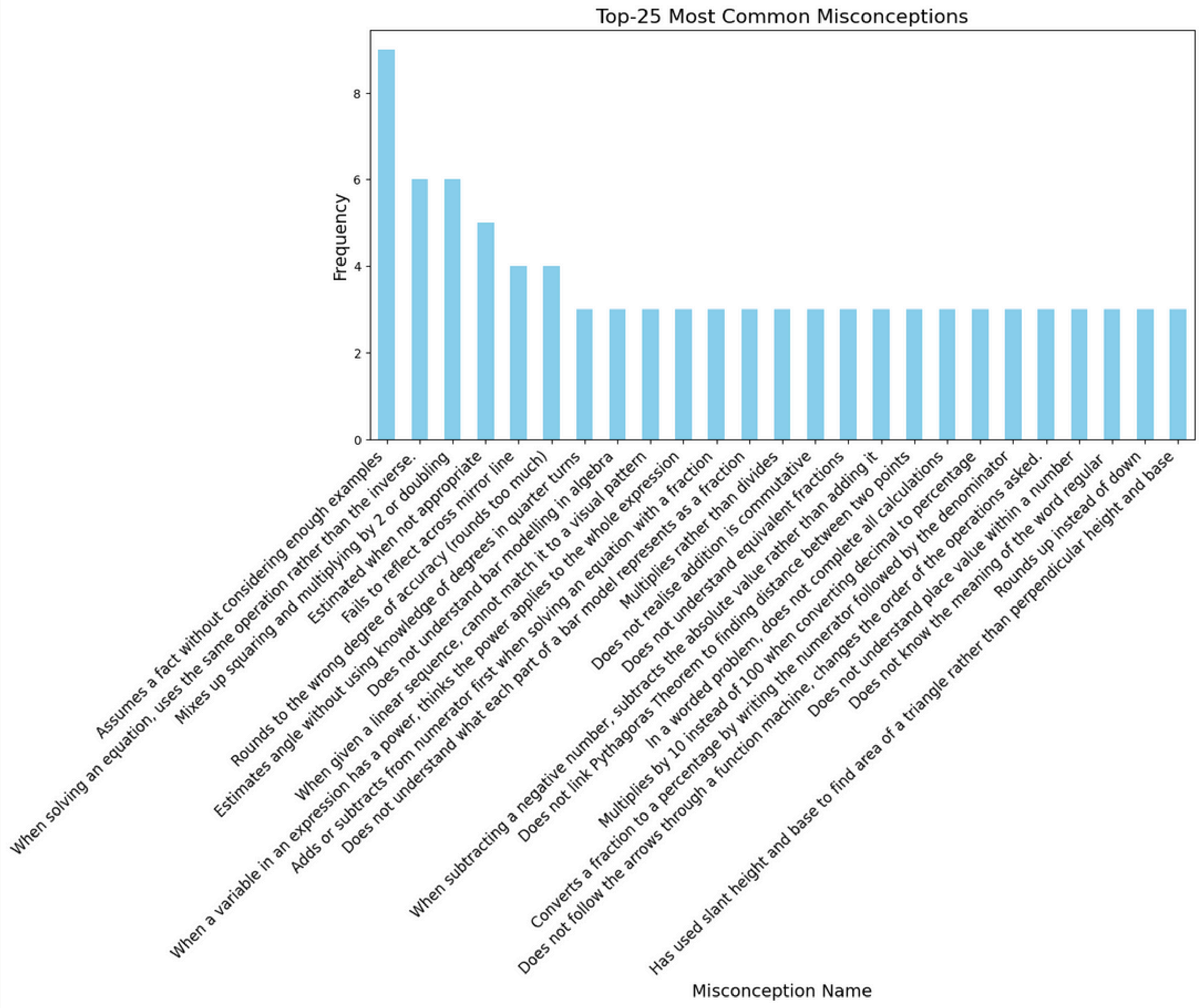
ConstructName: Calculate the square of a number,
Count: 14

SubjectName: Linear Equations,
Count: 53

About Dataset

misconception_mapping:

123	Does not know that angles in a triangle sum to 180 degrees
37	Uses dividing fractions method for multiplying fractions
246	Believes there are 100 degrees in a full turn



Evaluation

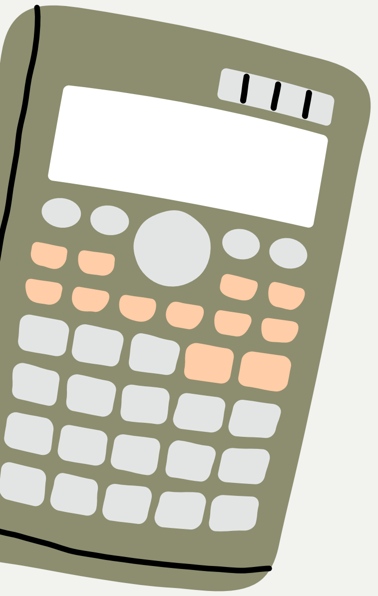
Equipment

U：觀測點的總數（即測試集的樣本數）。

n：每個樣本提交的預測數量（最多考慮前 25 個）。

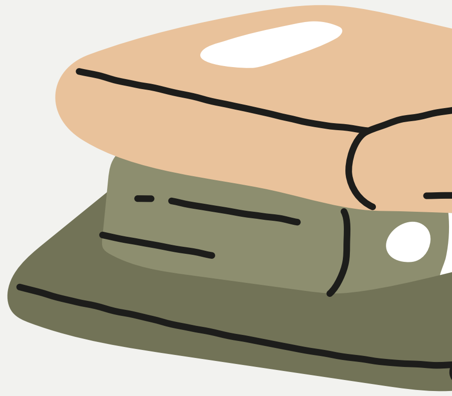
P(k)：在第 k 名的精度(排名越前數值越大)

rel(k)：相關性指示函數，當第 k 名的標籤是正確答案時，rel(k)=1，否則為 0。



Calculation formula:

$$\mathrm{MAP@25} = \frac{1}{U}\sum_{u=1}^{U}\sum_{k=1}^{min(n,25)} P(k)\times rel(k)$$



[A, B, C, D, E]

[A, A, A, A, A]

[A, B, A, C, A]

若A為正解，1、2、3皆為1

若B為正解，1、3皆為1/2

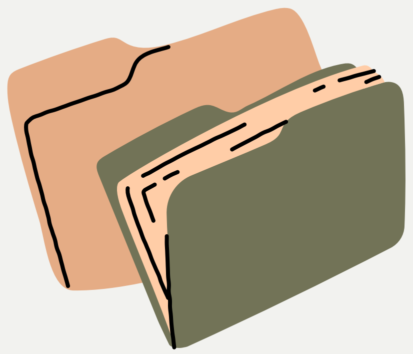
Eedi Qwen32B vllm with logits-processor-zoo

Qwen2.5-32B-Instruct-AWQ

- 採用 **Transformer 架構**，結合的技術如 RoPE（相對位置編碼）、SwiGLU（激活函數）、RMSNorm（正規化）以及 Attention QKV bias。
- 適合用於文本摘要、自然語言理解、程式碼生成與輔助
- 64 layers, with 40 attention heads for queries

Lora Finetune

- 對模型進行finetune，在資源較少的Kaggle十分有效
- 生成一個高效、記憶體友好的量化微調模型，準備進行推理或其他應用。



Code Review

Simply introduce code：Qwen14B_Retrieval_Qwen32B_logits-processor-zoo

Data Processing

- 生成每個選項（排除正確答案）作為錯誤選項的數據。
- 數據按照句子長度排序，確保短句和長句的填充處理一致。
- 利用 tokenizer 將文字轉為模型可處理的輸入格式。

First Retrieval

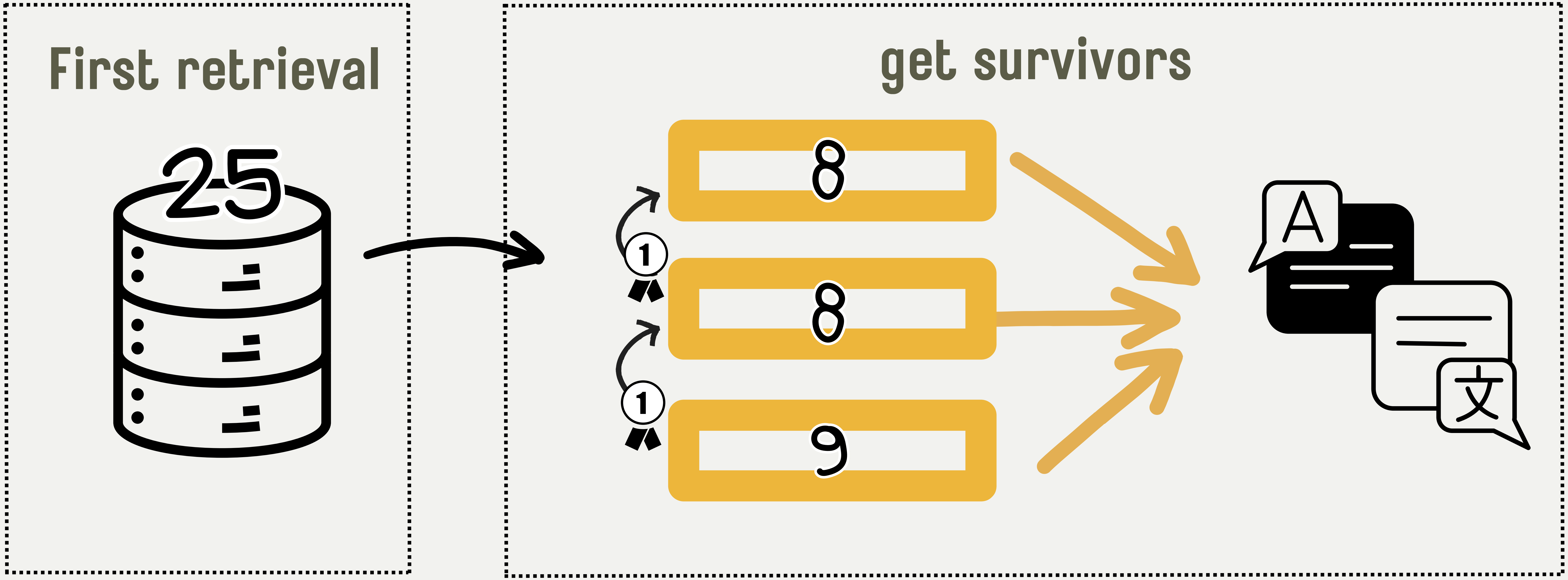
- Embedding： 使用"qwen2.5-14"模型透過inference函數將query_text進行embedding
- 使用NearestNeighbors找到最相似的25個嵌入向量對應的misconception

Mutiple Choice

- 使用 vLLM 調用 Qwen-32b-Instruct-AWQ 模型
- 逐輪篩選機制：3輪後選misconception，從候選中選擇最優項加入到下一階段，最終排序25個misconception
- 使用 LLM 根據生成的提示詞（text），通過多項選擇（Multiple Choice）生成每條數據的最佳選擇。

Code Review

Simply introduce code： Qwen14B_Retrieval_Qwen32B_logits-processor-zoo



Enhence

What we want to enhence code：Qwen14B_Retrieval_Qwen32B_logits-processor-zoo

Data Processing

- 將misconception的偏差值刪除
- 讓misconception有一致的格式

First Retrieval

- 增加 n_neighbors 的值
- 刪除數據中可能存在的異常值，它們會干擾最近鄰查詢的準確性。
- 使用不同於NearestNeighbors的初步篩選方法

Mutiple Choice

- 增加每輪篩選的survivors，或者增加輪數
- 給予排序高者更大的權重，使其較不易動搖
- 更改prompt來增強LLM的判斷

Code Review

Simply introduce code： Eedi Qwen-2.5 32B AWQ two-time retrieval

First-Time Retrieval → **LLM Reasoning** → **Second-Time Retrieval**

First-Time Retrieval

model: fine-tune model for Eedi dataset

1. Pre-processing: features = [” ConstructName”] + [”SubjectName”]
2. Embedding: features & misconceptions
3. 以semantic_search取出 top-100 misconceptions

LLM Reasoning

model: Qwen 2.5 32B-instruct-AWQ

1. input: prompt
2. output: [”llmMissconception”]
3. Post-processing: output

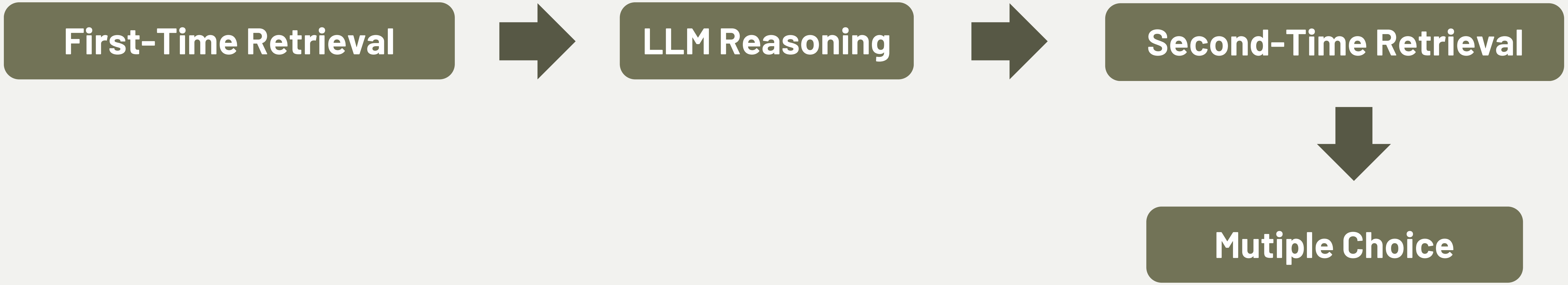
Second-Time Retrieval

model: fine-tune model for Eedi dataset

1. Pre-processing: features = prompt + [""llmMisconception_clean""]
2. Embedding: features
3. 以semantic_search取出top-25 misconceptions

Our Experiments

Combine two methods

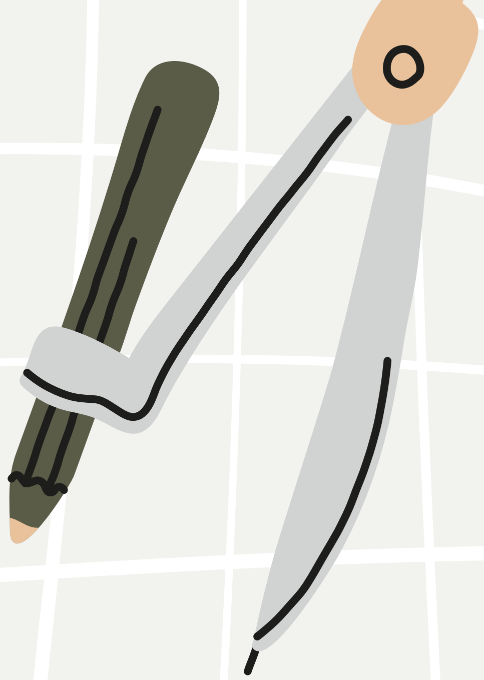


Eedi Qwen-2.5 32B AWQ two-time retrieval - Version 3

Notebook Threw Exception (after deadline) · 11h ago

Our Proposals

1. Using Augmented Data
2. Change Embedding Model
3. Two-time Reranking
4. Enhanced Retrieval



Thank you for listening.