

EEDI

MINING MISCONCEPTIONS IN MATHEMATICS

GVHD: TS. Nguyễn Tiến Huy

Trường Đại học Khoa học Tự nhiên - ĐHQG-HCM
Khoa Công nghệ thông tin

Ngày 18 tháng 1 năm 2025



Thành viên nhóm

MSSV	Tên thành viên
21120058	Phạm Nhật Duy
21120102	Nguyễn Trúc Nguyên
21120158	Trương Công Trung
21120279	Lê Trần Minh Khuê

Mục lục

- 1 Tổng quan dự án
- 2 Data Discovery
- 3 Data Preprocessing

Mục lục

- 4 Kiến trúc mô hình
 - Mô hình cơ sở
 - Tinh chỉnh mô hình
- 5 Triển khai mô hình
 - Ý tưởng
 - Giai đoạn 1: Truy xuất
 - Giai đoạn 2: Xếp hạng
- 6 Đánh giá mô hình
 - Độ đo
 - Kết quả đạt được

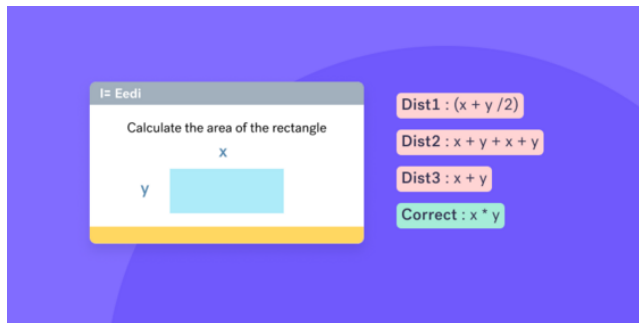
Tổng quan dự án

Giới thiệu chủ đề

Cuộc thi từ Eidi thách thức phát triển mô hình học máy xử lý ngôn ngữ tự nhiên, phân tích và dự đoán mối liên hệ giữa ngộ nhận và yếu tố gây nhiễu trong câu hỏi trắc nghiệm.

Giới thiệu chủ đề

Cuộc thi từ Eidi thách thức phát triển mô hình học máy xử lý ngôn ngữ tự nhiên, phân tích và dự đoán mối liên hệ giữa ngộ nhận và yếu tố gây nhiễu trong câu hỏi trắc nghiệm.



Giới thiệu chủ đề

Đặc biệt, đối với dữ liệu câu hỏi trắc nghiệm toán học, lựa chọn sai sẽ phản ánh sai lầm phổ biến trong tư duy. Mục tiêu chính là xây dựng mô hình vừa hiểu ý đồ của lựa chọn sai, vừa có thể khái quát hoá thành những ngộ nhận mới.

Giới thiệu chủ đề

Đặc biệt, đối với dữ liệu câu hỏi trắc nghiệm toán học, lựa chọn sai sẽ phản ánh sai lầm phổ biến trong tư duy. Mục tiêu chính là xây dựng mô hình vừa hiểu ý đồ của lựa chọn sai, vừa có thể khái quát hoá thành những ngộ nhận mới.

Problem:

$$\frac{6}{25} = ?$$

A. 0.625

B. 0.6

C. 6.25

D. 0.24

Correct answer: D

Misconception A: Converts a fraction to a decimal by placing all values of the fraction after the decimal point

Misconception B: Converts a fraction to a decimal by using only the numerator after the decimal point

Misconception C: Converts a fraction to a decimal by placing a decimal point between the numerator and denominator

Misconception D: No misconception

Construct Name: Convert fractions less than 1 to terminating decimals of 2 decimal places

Subject Name: Converting between Fractions and Decimals

Bộ dữ liệu

Bộ dữ liệu **Eedi** gồm 15 cột chứa thông tin về các câu hỏi trắc nghiệm và khái niệm liên quan, cùng với các câu trả lời.

Bộ dữ liệu

Bộ dữ liệu **Eedi** gồm 15 cột chứa thông tin về các câu hỏi trắc nghiệm và khái niệm liên quan, cùng với các câu trả lời.

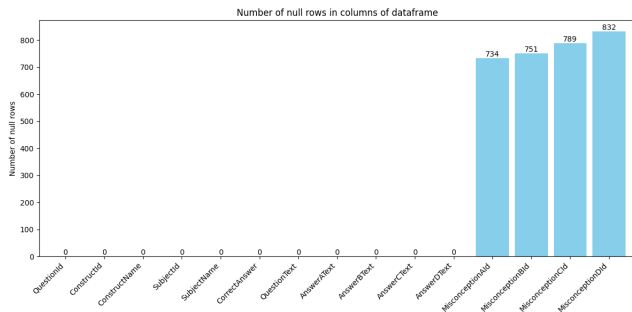
STT	Thuộc tính	Loại	Mô tả
1	QuestionId	int	Mã câu hỏi
2	ConstructId	int	Mã khái niệm
3	ConstructName	str	Khái niệm - Mức độ kiến thức chi tiết nhất liên quan tới câu hỏi
4	CorrectAnswer	char	Đáp án đúng (A, B, C, D)
5	SubjectId	int	Mã môn học
6	SubjectName	str	Tên môn học
7	QuestionText	str	Câu hỏi dạng văn bản
8	AnswerAText	str	Đáp án A dạng văn bản
9	AnswerBText	str	Đáp án B dạng văn bản
10	AnswerCText	str	Đáp án C dạng văn bản
11	AnswerDText	str	Đáp án D dạng văn bản
12	MisconceptionAId	int	Mã ngộ nhận tương ứng với đáp án A - nếu có
13	MisconceptionBId	int	Mã ngộ nhận tương ứng với đáp án B - nếu có
14	MisconceptionCId	int	Mã ngộ nhận tương ứng với đáp án C - nếu có
15	MisconceptionDId	int	Mã ngộ nhận tương ứng với đáp án D - nếu có

Mục tiêu dự án

- Phân tích dữ liệu để nhận diện đặc trưng quan trọng, mối quan hệ giữa câu hỏi, đáp án đúng, và đáp án gây nhiễu, cùng các mẫu quan niệm sai lầm phổ biến.
- Phát triển mô hình học máy dự đoán mối liên kết giữa quan niệm sai lầm và đáp án sai, tập trung nhận diện đặc trưng gây nhầm lẫn và khái quát hóa ngộ nhận mới.
- Đánh giá hiệu suất mô hình qua độ chính xác và khả năng tổng quát hóa trên các tập dữ liệu khác nhau.
- Tối ưu mô hình để đạt hiệu quả cao, đồng thời tìm các cải tiến tiềm năng cho ứng dụng thực tế.

Data Discovery

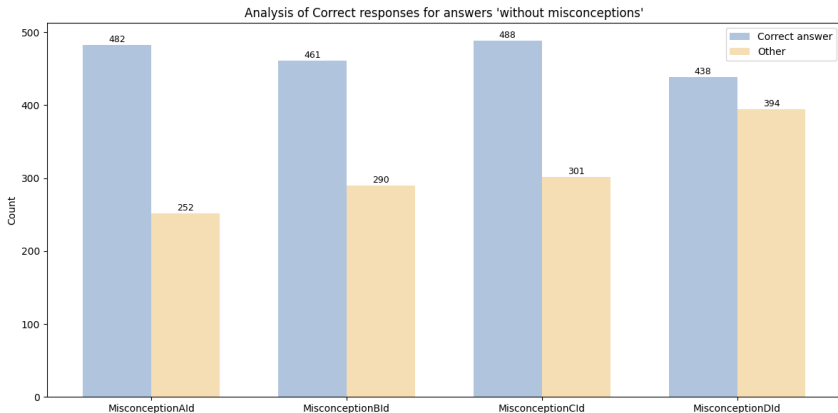
Số dòng null trong các cột của dataframe



Nhận xét

- Ta nhận thấy rằng chỉ có các cột Misconception bị NaN
- Các giá trị null này có thể là không có ngộ nhận, hoặc đây là đáp án đúng, hoặc là lỗi cơ bản, không liên quan đến bất kỳ ngộ nhận hay hiểu lầm nào về khái niệm.

Xem xét số lượng câu trả lời đúng của các câu trả lời 'không có ngộ nhận'



Phân tích chi tiết

Chỉ số	Số giá trị phân biệt	Giá trị xuất hiện nhiều nhất	Số lần xuất hiện nhiều nhất
ConstructName	757	Calculate the square of a number	14
SubjectName	163	Linear Equations	53
CorrectAnswer	4	C	488
QuestionText	1857	Which of the following pairs of function machines are correct?	4
AnswerAText	1219	Only Tom	93
AnswerBText	1230	Only Katie	109
AnswerCText	1222	Both Tom and Katie	158
AnswerDText	1184	Neither is correct	187

Phân tích chi tiết

Chỉ số	Số giá trị phân biệt	Giá trị xuất hiện nhiều nhất	Số lần xuất hiện nhiều nhất
ConstructName	757	Calculate the square of a number	14
SubjectName	163	Linear Equations	53
CorrectAnswer	4	C	488
QuestionText	1857	Which of the following pairs of function machines are correct?	4
AnswerAText	1219	Only Tom	93
AnswerBText	1230	Only Katie	109
AnswerCText	1222	Both Tom and Katie	158
AnswerDText	1184	Neither is correct	187

Nhận thấy rằng không có cột nào chỉ gồm một giá trị. Các khái niệm nền tảng đa dạng với 757 khái niệm riêng biệt. Ngoài ra, có tới 163 môn học/ phân môn khác nhau. Số lượng câu hỏi và câu trả lời rất đa dạng; tuy nhiên, có một số câu trả lời xuất hiện với tần suất khá cao, cần được xem xét kỹ lưỡng hơn.

Cột ConstructName

Index	ConstructName	Frequency
0	Calculate the square of a number	14
1	Solve two-step linear equations, with the variable on one side, with all positive integers	13
2	Factorise a quadratic expression in the form $x^2 + bx + c$	13
3	Use the order of operations to carry out calculations involving addition, subtraction, multiplication, and/or division	12
4	Identify the order of rotational symmetry of a shape	12
5	Multiply a single term over a bracket where the term on the outside is a number and the inside contains a linear expression	11
6	Calculate the range from a list of data	11
7	Solve one-step linear inequalities in one variable where the variable appears on one side of the equation	9
8	Recognise cube numbers	9
9	Solve quadratic equations using the quadratic formula where the coefficient of x^2 is not 1	9

Cột ConstructName

Index	ConstructName	Frequency
0	Calculate the square of a number	14
1	Solve two-step linear equations, with the variable on one side, with all positive integers	13
2	Factorise a quadratic expression in the form $x^2 + bx + c$	13
3	Use the order of operations to carry out calculations involving addition, subtraction, multiplication, and/or division	12
4	Identify the order of rotational symmetry of a shape	12
5	Multiply a single term over a bracket where the term on the outside is a number and the inside contains a linear expression	11
6	Calculate the range from a list of data	11
7	Solve one-step linear inequalities in one variable where the variable appears on one side of the equation	9
8	Recognise cube numbers	9
9	Solve quadratic equations using the quadratic formula where the coefficient of x^2 is not 1	9

Top 10 khái niệm nền tảng phổ biến nhất có mặt trong bộ dữ liệu là những khái niệm liên quan đến các chủ đề **đại số** và **hình học cơ bản**.

Cột SubjectName

Index	SubjectName	Frequency
0	Linear Equations	53
1	Linear Sequences (nth term)	44
2	BIDMAS	37
3	Quadratic Equations	36
4	Area of Simple Shapes	36
5	Writing Expressions	31
6	Place Value	31
7	Substitution into Formula	30
8	Multiplying and Dividing with Decimals	30
9	Function Machines	30
10	Reflection	28
11	Adding and Subtracting Fractions	28
12	Time	28
13	Mental Multiplication and Division	28
14	Basic Angle Facts (straight line, opposite, around a point, etc)	27
15	Averages (mean, median, mode) from a List of Data	26
16	Mental Addition and Subtraction	25
17	Real Life Graphs	25
18	Percentages of an Amount	24
19	Squares, Cubes, etc	24
...

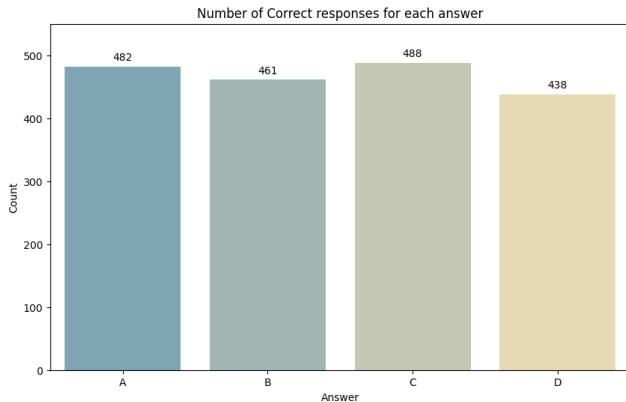
Phân tích cột SubjectName

Tần suất xuất hiện của từng SubjectName

Dữ liệu bao gồm nhiều nhóm chủ đề khác nhau: **đại số, hình học, số học, thống kê, và ứng dụng thực tế.**

- **Tần suất cao (30–50+):** Chủ đề nền tảng thường là những kiến thức cơ bản và quan trọng.
- **Tần suất trung bình (10–29):** Các chủ đề bổ trợ hoặc mở rộng (ví dụ như: “Real Life Graphs”, “Factorising into a Double Bracket”) thường xuất hiện đủ để cung cấp ngữ cảnh nhưng không phải trọng tâm chính.
- **Tần suất thấp (<10):** Các môn học chuyên sâu, mang tính nâng cao hoặc ít gặp trong thực tế (ví dụ như: “Surface Area of Prisms”, “Algebraic Proof”).

Cột CorrectAnswer



Nhận xét CorrectAnswer

Số lần các đáp án là câu trả lời đúng phân bố đồng đều.

Các ngộ nhận phổ biến

Hầu hết các ngộ nhận phổ biến (gặp ở nhiều câu hỏi) tập trung vào nhóm các quy tắc tính toán cơ bản:

Các ngộ nhận phổ biến

Hầu hết các ngộ nhận phổ biến (gặp ở nhiều câu hỏi) tập trung vào nhóm các quy tắc tính toán cơ bản: **Thứ tự ưu tiên của phép toán, và nhầm lẫn các phép toán**

MisconceptionId	MisconceptionName	Frequency
1214	When solving an equation, uses the same operation rather than the inverse.	54
2316	Mixes up squaring and multiplying by 2.	38
1507	Carries out operations from left to right regardless of priority order.	36
220	Only multiplies the first term in the expansion of a bracket.	22

Các ngộ nhận phổ biến

Hầu hết các ngộ nhận phổ biến (gặp ở nhiều câu hỏi) tập trung vào nhóm các quy tắc tính toán cơ bản: **Bản chất phép toán với số âm và phân số**

MisconceptionId	MisconceptionName	Frequency
1597	Believes multiplying two negatives gives a negative answer.	27
217	When adding fractions, adds the numerators and denominators.	16

Các ngộ nhận phổ biến

Hầu hết các ngộ nhận phổ biến (gặp ở nhiều câu hỏi) tập trung vào nhóm các quy tắc tính toán cơ bản: **Làm tròn số**

MisconceptionId	MisconceptionName	Frequency
1379	Rounds down instead of up.	43
2392, 1248	Rounds to the wrong degree of accuracy (too much: 27, too little: 22).	49

Các ngộ nhận phổ biến

Hầu hết các ngộ nhận phổ biến (gặp ở nhiều câu hỏi) tập trung vào nhóm các quy tắc tính toán cơ bản: **Nhận diện và sử dụng ký hiệu toán học**

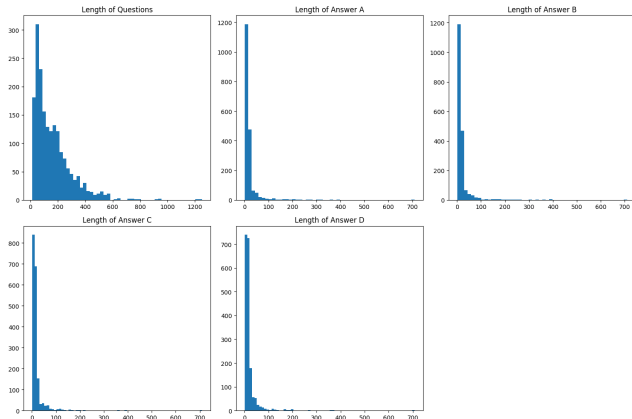
MisconceptionId	MisconceptionName	Frequency
1990	Fails to reflect across mirror line.	33
1880	Mixes up greater than and less than symbols.	32

Các ngộ nhận phổ biến

Hầu hết các ngộ nhận phổ biến (gặp ở nhiều câu hỏi) tập trung vào nhóm các quy tắc tính toán cơ bản.

Ngoài ra, còn có các ngộ nhận khác về tính toán số thực, hoặc các nhóm ngộ nhận khác nhưng chỉ **đặc thù** cho một số nhóm kiến thức/ lĩnh vực **nâng cao** hơn nên có tần suất xuất hiện thấp hơn.

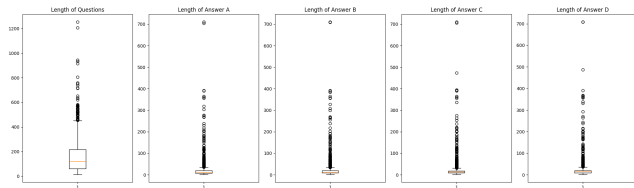
Phân tích độ dài câu hỏi



Nhận xét độ dài câu hỏi

Phân bố khá rộng, có cả câu hỏi ngắn và dài, nhưng tập trung chủ yếu ở mức ngắn (dưới 200 ký tự) và trung bình (từ 200 đến 600 ký tự).

Phân tích độ dài câu trả lời



Nhận xét độ dài câu trả lời

Có phần phân tán hơn nhiều. Tập trung nhiều nhất vẫn là nhóm câu trả lời ngắn, nhưng vẫn có một số câu trả lời khá dài hoặc dài so với phần còn lại.

Nhận xét phân phối dữ liệu

Kết luận

- Bộ dữ liệu có giá trị lớn và chứa nhiều thông tin hữu ích cho bài toán xác định mối quan hệ giữa các ngộ nhận và việc chọn đáp án sai.
- Bộ dữ liệu đa dạng về các phân môn và các phần kiến thức nền tảng trong toán học.
- Tuy nhiên, dữ liệu hiện tại hầu hết chỉ tập trung vào các bài toán với độ khó không quá cao. Hầu hết tập trung vào các phần kiến thức chung, chưa chuyên môn hoá. Vì vậy, mô hình sau khi huấn luyện có thể sẽ chạy tốt với các nhóm câu hỏi mức dễ, trung bình, khá, nhưng chưa thể giải quyết tốt được nhóm câu hỏi mức độ khó, hoặc thuộc các môn học nâng cao.

Data Preprocessing

Xử lý các ký tự dư thừa

Dữ liệu chứa các ký tự không cần thiết như ký tự xuống dòng, khoảng trắng thừa, link,... Nhóm tiến hành xử lý những ký tự này để làm sạch dữ liệu mà vẫn giữ nguyên định dạng LaTeX.

```
def clean_with_latex_preservation(x):  
    # Remove links  
    x = re.sub("http\\w+", " ", x)  
    # Remove newline, tab, and carriage return characters  
    x = re.sub(r"[\n\t\r]+", " ", x)  
    # Remove extra spaces  
    x = re.sub(r"[ ]{2,}", " ", x)  
    # Preserve LaTeX formatting by removing unintended whitespace after backslashes  
    x = re.sub(r"(?!\\)\s+", r"\\", x)  
    # Trim leading and trailing spaces  
    x = x.strip()  
    return x
```

Chuẩn hoá dữ liệu đầu vào

Prompt cho mô hình sẽ có đầy đủ thông tin gồm nội dung câu hỏi, câu trả lời đúng, câu trả lời sai mà học sinh đã chọn.

```
TEMPLATE_INPUT = '{QUESTION}\n  
Correct answer: {CORRECT_ANSWER}\n  
Student wrong answer: {STUDENT_WRONG_ANSWER}'
```

◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ↺ 🔍 ↻ 29/59

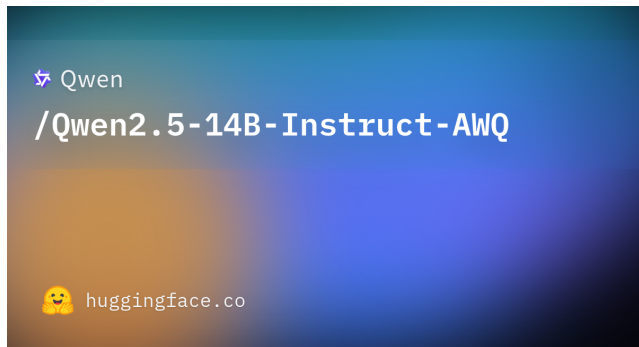
- 1 Tổng quan dự án
- 2 Data Discovery
- 3 Data Preprocessing
- 4 Kiến trúc mô hình**
 - Mô hình cơ sở
 - Tinh chỉnh mô hình
- 5 Triển khai mô hình
 - Ý tưởng
 - Giai đoạn 1: Truy xuất
 - Giai đoạn 2: Xếp hạng
- 6 Đánh giá mô hình
 - Độ đo
 - Kết quả đạt được

Lựa chọn mô hình

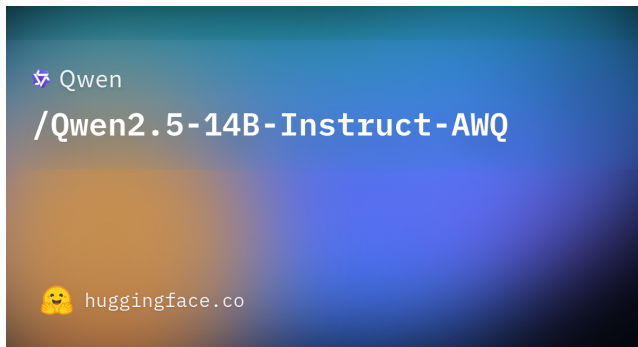
Trong quá trình xây dựng và áp dụng giải pháp để xác định ngộ nhận trong các bài toán toán học, việc lựa chọn mô hình ngôn ngữ lớn là yếu tố then chốt nhằm đảm bảo hiệu quả và độ chính xác.

Lựa chọn mô hình

Trong quá trình xây dựng và áp dụng giải pháp để xác định ngộ nhận trong các bài toán toán học, việc lựa chọn mô hình ngôn ngữ lớn là yếu tố then chốt nhằm đảm bảo hiệu quả và độ chính xác.



Mô hình cơ sở



Mô hình Qwen2.5-14B được đánh giá cao nhờ khả năng xử lý ngôn ngữ tự nhiên ở mức độ nâng cao, đặc biệt trong các bài toán yêu cầu phân tích và suy luận sâu.

- 1 Tổng quan dự án
- 2 Data Discovery
- 3 Data Preprocessing
- 4 Kiến trúc mô hình**
 - Mô hình cơ sở
 - Tinh chỉnh mô hình
- 5 Triển khai mô hình
 - Ý tưởng
 - Giai đoạn 1: Truy xuất
 - Giai đoạn 2: Xếp hạng
- 6 Đánh giá mô hình
 - Độ đo
 - Kết quả đạt được

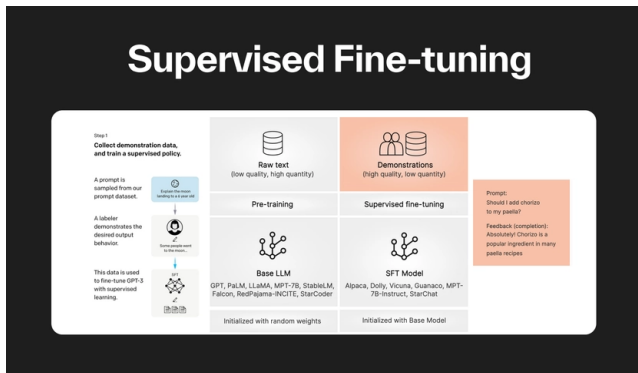
Tinh chỉnh mô hình

Để nâng cao hiệu suất của Qwen2.5-14B khi áp dụng vào bài toán xác định ngộ nhận toán học, nhóm đã quyết định sử dụng bộ dữ liệu Eedi trong giai đoạn tiền huấn luyện.

Quy trình này bao gồm hai thành phần chính: huấn luyện có giám sát (**SFT**) và ứng dụng kỹ thuật **AWQ**.

Supervised Fine-Tuning - SFT

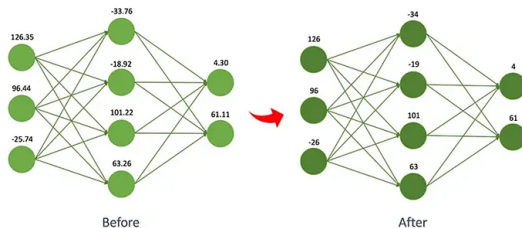
Dữ liệu sẽ được định dạng theo cấu trúc: [Chủ đề môn học] → [Câu hỏi] → [Phản hồi của học sinh] → [Mô tả ngộ nhận]. Cấu trúc này giúp mô hình học cách nhận diện ngộ nhận dựa trên ngữ cảnh.



Activation-aware Weight Quantization - AWQ

AWQ giúp biểu diễn trọng số của mô hình bằng các số nguyên 4-bit hoặc 8-bit thay vì sử dụng các số thực 16-bit hoặc 32-bit truyền thống.

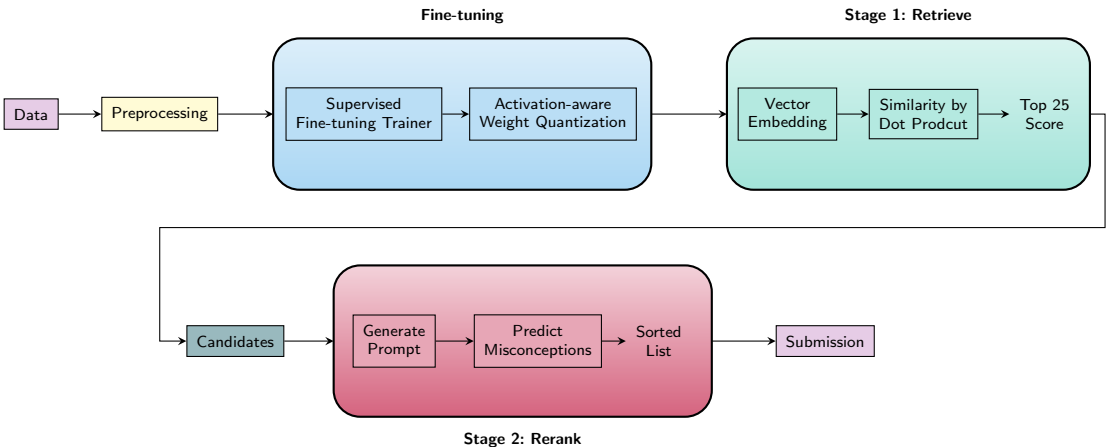
Quantize LLMs Using AWQ



◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ↺ 🔍 ↻ 37/59

- 1 Tổng quan dự án
- 2 Data Discovery
- 3 Data Preprocessing
- 4 Kiến trúc mô hình
 - Mô hình cơ sở
 - Tinh chỉnh mô hình
- 5 Triển khai mô hình**
 - Ý tưởng
 - Giai đoạn 1: Truy xuất
 - Giai đoạn 2: Xếp hạng
- 6 Đánh giá mô hình
 - Độ đo
 - Kết quả đạt được

Ý tưởng



- 1 Tổng quan dự án
- 2 Data Discovery
- 3 Data Preprocessing
- 4 Kiến trúc mô hình
 - Mô hình cơ sở
 - Tinh chỉnh mô hình
- 5 Triển khai mô hình**
 - Ý tưởng
 - **Giai đoạn 1: Truy xuất**
 - Giai đoạn 2: Xếp hạng
- 6 Đánh giá mô hình
 - Độ đo
 - Kết quả đạt được

Tạo truy vấn

Đầu tiên, tạo các truy vấn chi tiết từ dữ liệu đầu vào cho mô hình bằng cách kết hợp giữa mô tả nhiệm vụ (`task_description`) và thông tin cụ thể từ dữ liệu đầu vào (`query`).

```
def get_detailed_instruct(task_description: str, query: str) -> str:
    return f'<instruct>{task_description}\n<query>{query}'

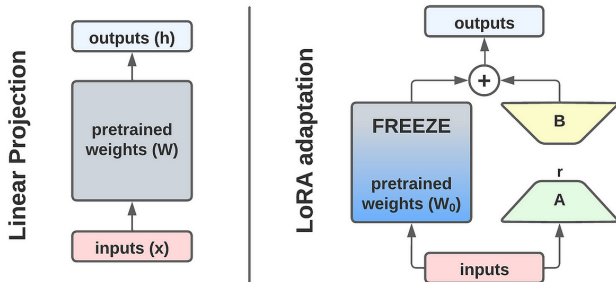
task = "Given a math multiple-choice problem with a student's wrong
        answer, retrieve the math misconceptions"
queries = [
    get_detailed_instruct(task, q) for q in df_input['Prompt']
]
```


Tạo truy vấn

Token hóa văn bản, tính toán trạng thái ẩn (hidden states) và trích xuất vector đại diện từ trạng thái cuối cùng của mô hình \Rightarrow **Vector Embedding**.

```
batch_embeddings = last_token_pool(outputs.last_hidden_state ,  
                                   batch_dict["attention_mask"])  
batch_embeddings = F.normalize(batch_embeddings , p=2, dim=1)
```

LoRA



Áp dụng kỹ thuật LoRA

- Giảm số lượng tham số cần cập nhật, từ đó giảm tải bộ nhớ và chi phí tính toán.
- LoRA thêm các tham số phụ trợ để huấn luyện, đảm bảo tính linh hoạt và dễ áp dụng cho các mô hình đã được huấn luyện trước.

Xác định mức độ tương đồng

Mức độ tương đồng ngữ nghĩa giữa các cặp truy vấn - ngộ nhận được tính bằng tích vô hướng giữa vector nhúng của truy vấn (`query_embeddings`) và các vector nhúng của tài liệu (`doc_embeddings`).

```
scores = query_embeddings @ doc_embeddings.T
```

Top 25 ngộ nhận có điểm tương đồng cao nhất \Rightarrow **Candidates**.

- 1 Tổng quan dự án
- 2 Data Discovery
- 3 Data Preprocessing
- 4 Kiến trúc mô hình
 - Mô hình cơ sở
 - Tinh chỉnh mô hình
- 5 Triển khai mô hình**
 - Ý tưởng
 - Giai đoạn 1: Truy xuất
 - Giai đoạn 2: Xếp hạng**
- 6 Đánh giá mô hình
 - Độ đo
 - Kết quả đạt được

Tạo Prompt chi tiết

```
PROMPT = """Here is a question about {ConstructName}({SubjectName}).  
Question: {Question}  
Correct Answer: {CorrectAnswer}  
Incorrect Answer: {IncorrectAnswer}
```

```
You are a Mathematics teacher. Your task is to reason and identify the  
misconception behind the Incorrect Answer with the Question.  
Answer concisely what misconception it is to lead to getting the  
Incorrect Answer.
```

```
Pick the correct misconception number from the below:
```

```
{Retrieval}  
"""
```

Quy trình xếp hạng và dự đoán ngộ nhận

Xếp hạng ngộ nhận qua từng vòng lặp:

- Tạo danh sách ngộ nhận tiềm năng ($c_indices$) dựa trên điểm tương đồng từ các vòng trước đó.

Quy trình xếp hạng và dự đoán ngộ nhận

Xếp hạng ngộ nhận qua từng vòng lặp:

- Tạo danh sách ngộ nhận tiềm năng ($c_indices$) dựa trên điểm tương đồng từ các vòng trước đó.
- Ghép danh sách này vào prompt để mô hình ngôn ngữ phân tích và lựa chọn.

Quy trình xếp hạng và dự đoán ngộ nhận

Dùng LLM để chọn ngộ nhận:

- Áp dụng prompt cho từng câu hỏi trong tập dữ liệu.

Quy trình xếp hạng và dự đoán ngộ nhận

Dùng LLM để chọn ngộ nhận:

- Áp dụng prompt cho từng câu hỏi trong tập dữ liệu.
- Sử dụng mô hình để dự đoán ngộ nhận phù hợp nhất thông qua các tham số:
 - Chế độ lấy mẫu ($\text{top_k}=1$, $\text{temperature}=0$) đảm bảo kết quả có độ chính xác cao.
 - **Logits processor:** Giới hạn lựa chọn các ngộ nhận trong danh sách đã cung cấp (`MultipleChoiceLogitsProcessor`).

Quy trình xếp hạng và dự đoán ngộ nhận

Lưu kết quả và tối ưu qua các vòng lặp:

- Mỗi vòng, LLM chọn ngộ nhận tốt nhất từ danh sách.

Quy trình xếp hạng và dự đoán ngộ nhận

Lưu kết quả và tối ưu qua các vòng lặp:

- Mỗi vòng, LLM chọn ngộ nhận tốt nhất từ danh sách.
- Các ngộ nhận được cập nhật, giảm thiểu nhiễu và tăng độ chính xác.

Quy trình xếp hạng và dự đoán ngộ nhận

Kết quả phân tích

Here is a question about: *Use the order of operations to carry out calculations involving powers (BIDMAS).*

Question:

$$3 \times 2 + 4 - 5$$

Where do the brackets need to go to make the answer equal 13?

Correct Answer:

$$3 \times (2 + 4) - 5$$

Incorrect Answer:

$$3 \times 2 + (4 - 5)$$

Quy trình xếp hạng và dự đoán ngộ nhận

Kết quả phân tích

Task: You are a Mathematics teacher. Your task is to reason and identify the misconception behind the Incorrect Answer with the Question.

Answer concisely what misconception it is to lead to getting the Incorrect Answer.

Pick the correct misconception number from the below:

- 1** Performs subtraction right to left if priority order means doing a calculation to the right first
- 2** Believes order of operations does not affect the answer to a calculation
- 3** Performs addition ahead of any other operation
- 4** Answers order of operations questions with brackets as if the brackets are not there
- 5** Performs addition ahead of subtraction
- 6** Does not interpret the correct order of operations from a worded problem
- 7** Confuses the order of operations, believes subtraction comes before multiplication
- 8** Believes that the order of a worded calculation should be changed to follow BIDMAS
- 9** Has not realised that the answer may be changed by the insertion of brackets

Đánh giá mô hình

- 1 Tổng quan dự án
- 2 Data Discovery
- 3 Data Preprocessing
- 4 Kiến trúc mô hình
 - Mô hình cơ sở
 - Tinh chỉnh mô hình
- 5 Triển khai mô hình
 - Ý tưởng
 - Giai đoạn 1: Truy xuất
 - Giai đoạn 2: Xếp hạng
- 6 Đánh giá mô hình**
 - **Độ đo**
 - Kết quả đạt được

Average Precision (AP)

AP được tính cho một cặp nhãn thực tế và dự đoán, giới hạn trong *top-k* kết quả.

$$AP = \frac{1}{\min(\text{len}(\text{actual}), k)} \sum_{i=1}^k \frac{\text{num_hits}}{i}$$

Mean Average Precision (MAP)

MAP là trung bình của AP trên toàn bộ danh sách nhãn thực tế và dự đoán.

$$MAP = \frac{1}{N} \sum_{i=1}^N AP_i$$

- 1 Tổng quan dự án
- 2 Data Discovery
- 3 Data Preprocessing
- 4 Kiến trúc mô hình
 - Mô hình cơ sở
 - Tinh chỉnh mô hình
- 5 Triển khai mô hình
 - Ý tưởng
 - Giai đoạn 1: Truy xuất
 - Giai đoạn 2: Xếp hạng
- 6 Đánh giá mô hình**
 - Độ đo
 - Kết quả đạt được

Kết quả Kaggle

Điểm số do hệ thống chấm của cuộc thi ghi nhận là minh chứng cho sự chính xác và khả năng ứng dụng thực tế của giải pháp.

Leaderboard

Raw Data

Refresh

Q HCMUS

Public







Private

The private leaderboard is calculated with approximately 72% of the test data. This competition has completed. This leaderboard reflects the final standings.



Prize Winners

#	△	Team	Members	Score	Entries	Last	Solution
114	▲ 15	HCMUS - Senior Students		0.47056	346	1mo	
152	▲ 21	HCMUS-2425-1		0.45878	27	1mo	

Tài liệu tham khảo I

-  Z. Jiang, “Llms can find mathematical reasoning mistakes by pedagogical chain-of-thought,” 2024.
-  B. C. Sadihin, “Mining misconceptions in mathematics,” 2024.
-  F. M. Aguilera, “Sfttrainer: A comprehensive exploration of its concept, advantages, limitations, history, and applications,” 2024.
-  A. Vo, “Retrieval with qwen-14b-awq,” 2024.
-  M. Dawood, “Understanding the competition : Eda and overview,” 2024.
-  M. Prata, “Misconceptions distractors in maths,” 2024.

Tài liệu tham khảo II

-  J. Kiran, "Qwen2.5-32b-awq for identifying likely misconceptions in reranking," 2024.
-  E. Albayrak, "Qwen instruct awq - vllm tutorial," 2024.