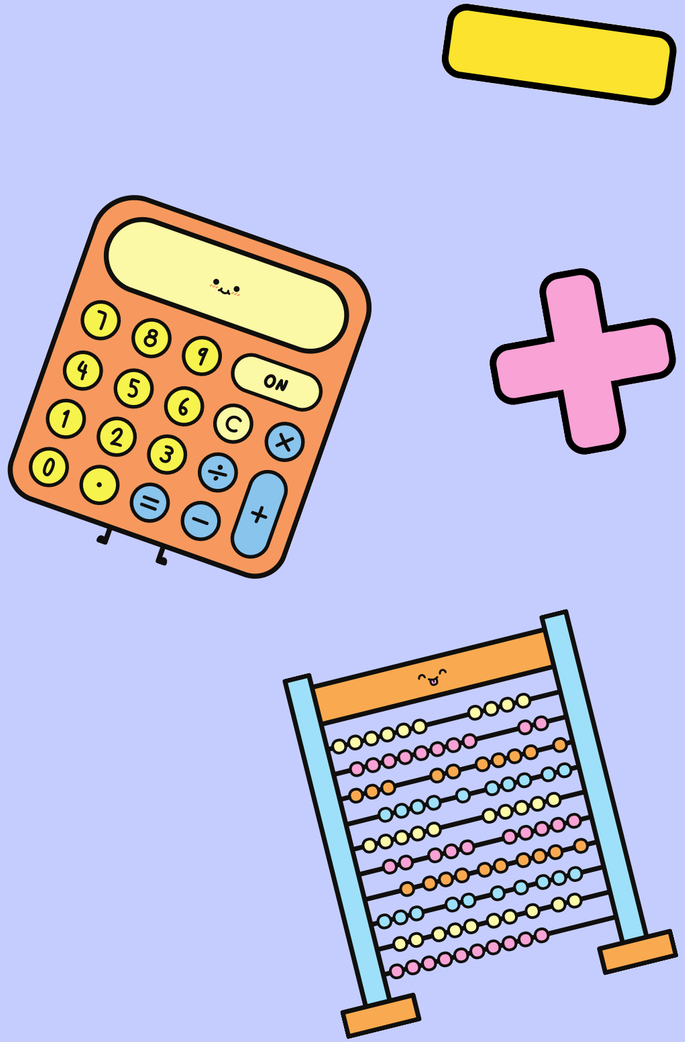
Predict affinity between misconceptions and incorrect answers (distractors) in multiple-choice questions in Mathematics

Prince & Vanshith

DSDA 385

Content

- Dataset
- Model
- Loss function
- Results



Dataset

- Train.csv
- Test.csv
- Misconception Mapping.csv

Train Dataset:

	QuestionId	ConstructId	ConstructName
0	0	856	Use the order of operations to carry out calcu...
1	1	1612	Simplify an algebraic fraction by factorising ...
2	2	2774	Calculate the range from a list of data
3	3	2377	Recall and use the intersecting diagonals prop...
4	4	3387	Substitute positive integer values into formul...

	SubjectId	SubjectName	CorrectAnswer
0	33	BIDMAS	A
1	1077	Simplifying Algebraic Fractions	D
2	339	Range and Interquartile Range from a List of Data	B
3	88	Properties of Quadrilaterals	C
4	67	Substitution into Formula	A

	QuestionText	AnswerAText
0	\[\n3 \times 2+4-5\n\]\nWhere do the brackets ...	\(3 \times(2+4)-5 \)
1	Simplify the following, if possible: \(\frac{...	\(m+1 \)
2	Tom and Katie are discussing the \(5 \) plant...	Only\nTom
3	The angles highlighted on this rectangle with ...	acute
4	The equation \(f=3 r^{2}+3 \) is used to find...	\(30 \)

	AnswerBText	AnswerCText	AnswerDText
0	\(3 \times 2+(4-5) \)	\(3 \times(2+4-5) \)	Does not need brackets
1	\(m+2 \)	\(m-1 \)	Does not simplify
2	Only\nKatie	Both Tom and Katie	Neither is correct
3	obtuse	\(90^{\circ} \)	Not enough information
4	\(27 \)	\(51 \)	\(24 \)

	MisconceptionAId	MisconceptionBId	MisconceptionCId	MisconceptionDId
0	NaN	NaN	NaN	1672.0
1	2142.0	143.0	2142.0	NaN
2	1287.0	NaN	1287.0	1073.0
3	1180.0	1180.0	NaN	1180.0
4	NaN	NaN	NaN	1818.0

Misconception Mapping:

	MisconceptionId	MisconceptionName
0	0	Does not know that angles in a triangle sum to...
1	1	Uses dividing fractions method for multiplying...
2	2	Believes there are 100 degrees in a full turn
3	3	Thinks a quadratic without a non variable term...
4	4	Believes addition of terms and powers of terms...

Train Dataset Info:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1869 entries, 0 to 1868
Data columns (total 15 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   QuestionId         1869 non-null   int64
 1   ConstructId        1869 non-null   int64
 2   ConstructName      1869 non-null   object
 3   SubjectId          1869 non-null   int64
 4   SubjectName        1869 non-null   object
 5   CorrectAnswer      1869 non-null   object
 6   QuestionText       1869 non-null   object
 7   AnswerAText        1869 non-null   object
 8   AnswerBText        1869 non-null   object
 9   AnswerCText        1869 non-null   object
 10  AnswerDText        1869 non-null   object
 11  MisconceptionAId   1135 non-null   float64
 12  MisconceptionBId   1118 non-null   float64
 13  MisconceptionCId   1080 non-null   float64
 14  MisconceptionDId   1037 non-null   float64
dtypes: float64(4), int64(3), object(8)
memory usage: 219.1+ KB
None

Total Unique Misconceptions: 1604
```

Construct Name: Simplify an algebraic fraction by factorising the numerator

Subject Name: Simplifying Algebraic Fractions

Problem:

Simplify the following, if possible: $\frac{m^2+2m-3}{m-3}$

A. $m+1$

B. $m+2$

C. $m-1$

D. Does not simplify

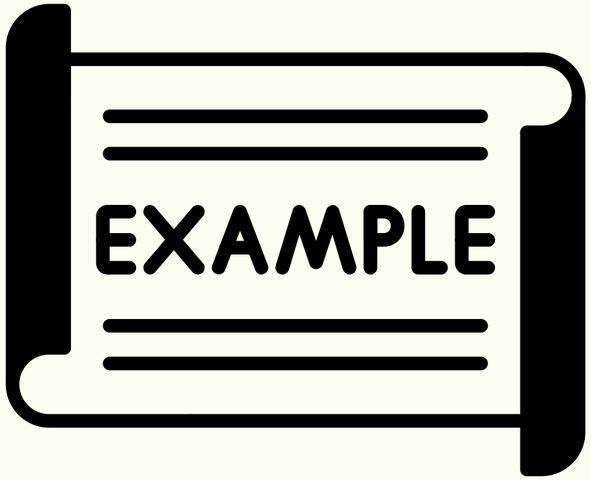
Correct answer: D

Misconception A: Does not know that to factorise a quadratic expression, to find two numbers that add to give the coefficient of the x term, and multiply to give the non variable term

Misconception B: Thinks that when you cancel identical terms from the numerator and denominator, they just disappear

Misconception C: Does not know that to factorise a quadratic expression, to find two numbers that add to give the coefficient of the x term, and multiply to give the non variable term

Misconception D: No misconception/NaN



Misconceptions are mapped as IDs and have separate mapping database.
1604 Unique Misconceptions labels

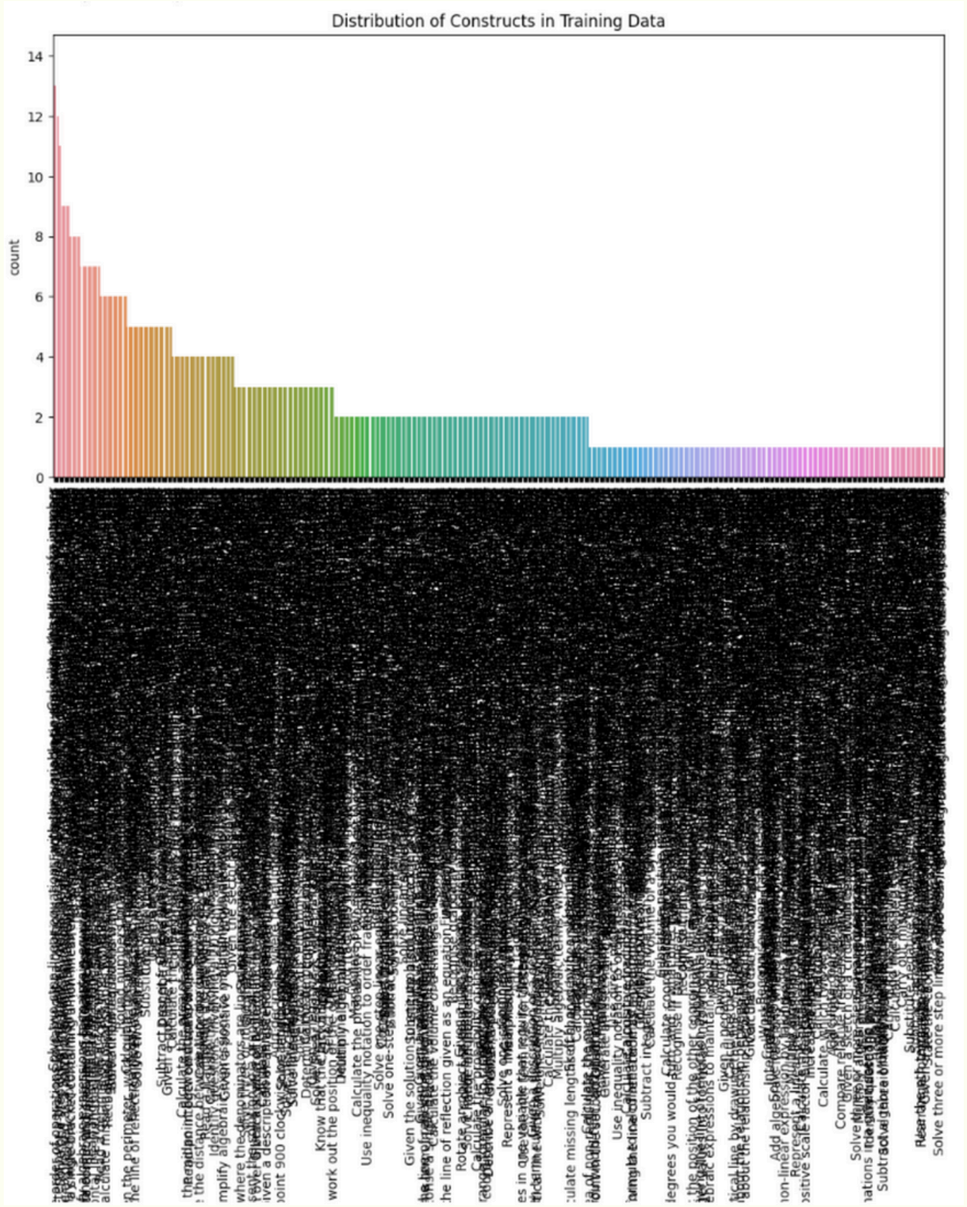
Model

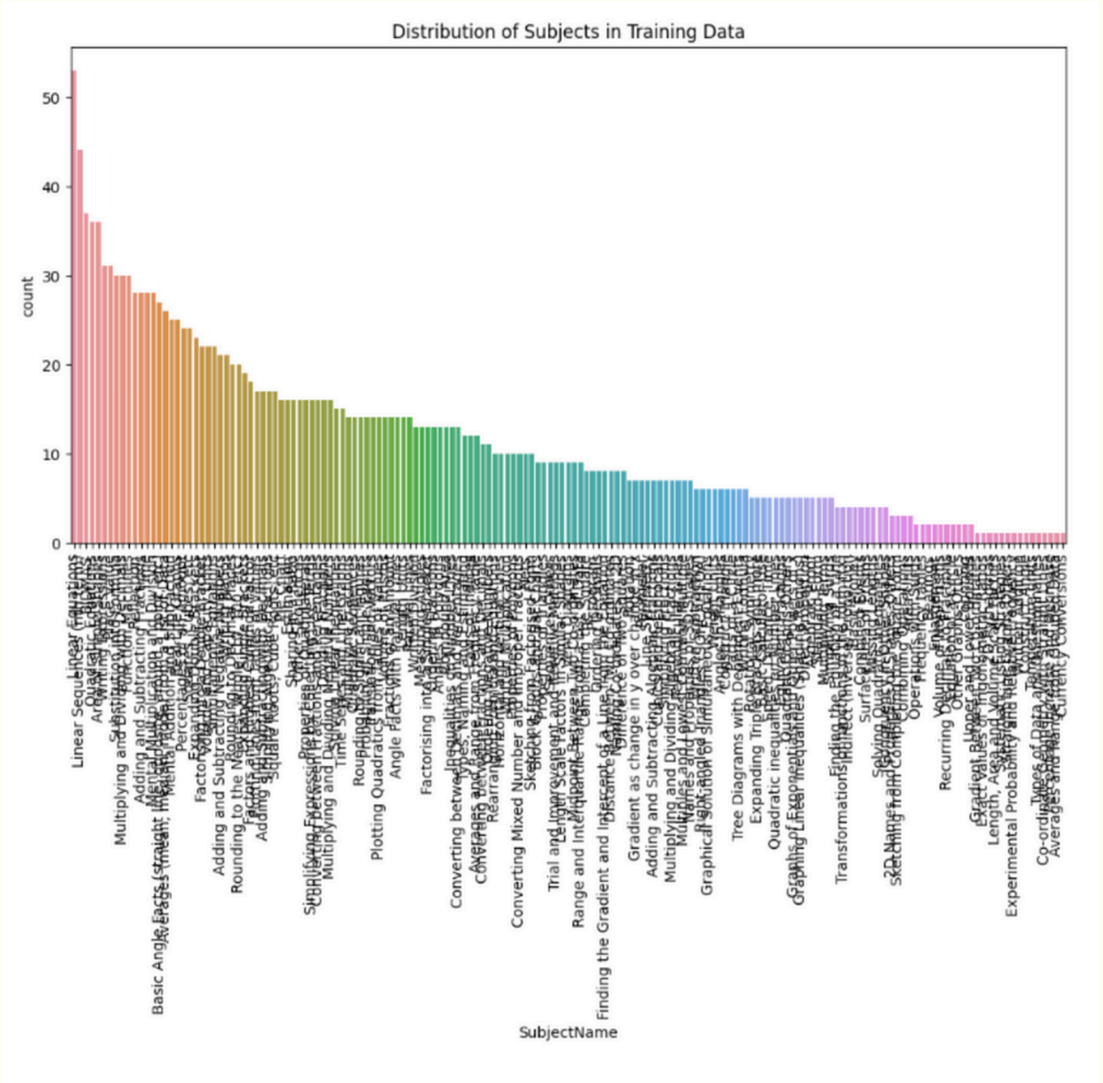
Regression, NN Prediction?

Won't be possible to EXECUTE as most text variables like Subject, Construct, Misconceptions, Question are Unique and the textual understanding is significant not the variable itself.

Large Language Model and NLP?

Given the diversity and uniqueness of misconceptions, constructs, and subjects in this dataset, a language-based model such as a transformer-based architecture is more appropriate for capturing the underlying relationships and understanding the semantics of the questions and misconceptions.





Subject Names are Significant to use in our Model as there's Repetition

Before Jumping to Model

Why are we doing this ?

Why and Impact ?

The goal is to create a model that not only aligns with known misconceptions but also generalizes to new, emerging misconceptions. Such a model would assist human labelers in accurately selecting suitable misconceptions from both existing and newly identified options.

It could help improve the understanding and management of misconceptions, enhancing the educational experience for both students and teachers.

It could also be a way to train Generative Models on how to associate distractors to Misconceptions which can assist in teaching when being used.

Data Preprocessing

Test Data.csv → Query Text Creation and Eliminate Correct Answer and Have each row of Incorrect Answer. Task is to predict the Misconceptions associated with each Incorrect Answer Multiple Choice in a Question

	query_text	QuestionId_Answer	ConstructName	SubjectName	QuestionText	correct_answer	incorrect_answer
0	### SubjectName: BIDMAS\n### ConstructName: Us...	1869_B	Use the order of operations to carry out calcu...	BIDMAS	$\n3 \times 2+4-5\n$ \nWhere do the brackets ...	$3 \times (2+4)-5$	$3 \times 2+(4-5)$
1	### SubjectName: BIDMAS\n### ConstructName: Us...	1869_C	Use the order of operations to carry out calcu...	BIDMAS	$\n3 \times 2+4-5\n$ \nWhere do the brackets ...	$3 \times (2+4)-5$	$3 \times (2+4-5)$
2	### SubjectName: BIDMAS\n### ConstructName: Us...	1869_D	Use the order of operations to carry out calcu...	BIDMAS	$\n3 \times 2+4-5\n$ \nWhere do the brackets ...	$3 \times (2+4)-5$	Does not need brackets
3	### SubjectName: Simplifying Algebraic Fractio...	1870_A	Simplify an algebraic fraction by factorising ...	Simplifying Algebraic Fractions	Simplify the following, if possible: \(\frac{...	Does not simplify	$m+1$
4	### SubjectName: Simplifying Algebraic Fractio...	1870_B	Simplify an algebraic fraction by factorising ...	Simplifying Algebraic Fractions	Simplify the following, if possible: \(\frac{...	Does not simplify	$m+2$
5	### SubjectName: Simplifying Algebraic Fractio...	1870_C	Simplify an algebraic fraction by factorising ...	Simplifying Algebraic Fractions	Simplify the following, if possible: \(\frac{...	Does not simplify	$m-1$
6	### SubjectName: Range and Interquartile Range...	1871_A	Calculate the range from a list of data	Range and Interquartile Range from a List of Data	Tom and Katie are discussing the 5 plant...	Only\nKatie	Only\nTom
7	### SubjectName: Range and Interquartile Range...	1871_C	Calculate the range from a list of data	Range and Interquartile Range from a List of Data	Tom and Katie are discussing the 5 plant...	Only\nKatie	Both Tom and Katie
8	### SubjectName: Range and Interquartile Range...	1871_D	Calculate the range from a list of data	Range and Interquartile Range from a List of Data	Tom and Katie are discussing the 5 plant...	Only\nKatie	Neither is correct

```python
def process_row(row):
    rows = []
    for option in ["A", "B", "C", "D"]:
        if option == row['CorrectAnswer']:
            continue
        correct_answer = row[f"Answer{row['CorrectAnswer']}Text"]
        query_text = (f"### SubjectName: {row['SubjectName']}\n"
                      f"### ConstructName: {row['ConstructName']}\n"
                      f"### Question: {row['QuestionText']}\n"
                      f"### Correct Answer: {correct_answer}\n"
                      f"### Misconception Incorrect answer: {option}.{
```

Approach

Only Test Data Preprocessing

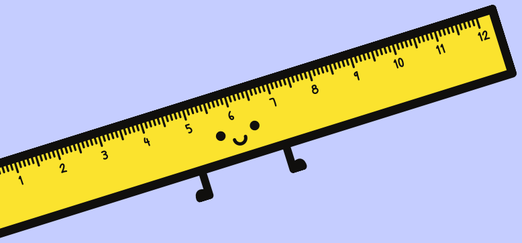
Retrieval Model Qwen 14B

Misconception Prediction using Test Data

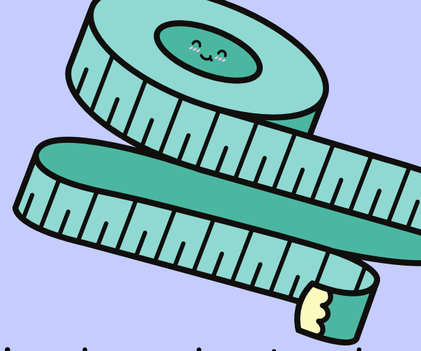
Misconceptions Embeddings

K-nearest Mean 25 cluster matches

Qwen 32B + logits-processor-zoo



created by Alibaba Cloud



Qwen Models

a large-scale Transformer-based language model. Its architecture follows the standard design of attention-based networks:

- Multi-Head Attention
- Layer Normalization
- Feedforward Networks

We further enhance it using LoRA (Low-Rank Adaptation), a parameter-efficient fine-tuning method. LoRA modifies specific attention layers (like q_proj and v_proj) by learning compact, low-rank updates, keeping the rest of the model frozen.

The model generates semantic embeddings, numerical representations of the input text, in a high-dimensional vector space. These embeddings are normalized using cosine similarity, ensuring that text with similar meanings appears closer in the embedding space. The last-token pooling strategy aggregates the representation of sequences based on their actual lengths, ensuring contextually rich embeddings.

Contrastive Loss

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{\exp(\text{sim}(\mathbf{q}_i, \mathbf{c}_i)/\tau)}{\sum_{j=1}^{N}\exp(\text{sim}(\mathbf{q}_i, \mathbf{c}_j)/\tau)}$$

- sim(qi,ci): Cosine similarity between paired embeddings.
- τ: Temperature parameter controlling sharpness. This approach ensures better generalization for unseen data.

a function designed to pull embeddings of correct pairs (e.g., incorrect answers and their corresponding misconceptions) closer and push embeddings of incorrect pairs further apart in the embedding space.

Logits processor zoo

LLM Zoo of Tools by Nvidia.

- We are Using MultipleChoiceLogitsProcessor.
- It guides the model to generate responses that correspond to the provided choices, ensuring accurate and relevant answers.

Imagine you ask:
"What is 2 + 2?"
Choices - 3, 4, 5

Without a logits processor, the model could output anything, like "6" or "bananaz." The MultipleChoiceLogitsProcessor adjusts the model's "thinking" by:

- Highlighting the tokens that match the answers (e.g., tokens for "3," "4," and "5").
- Penalizing all other tokens in its vocabulary.

Now, the model is nudged to choose one of the valid answers (3, 4, or 5).

It does this in two ways

- Filtering or boosting logits: It increases the probabilities of tokens related to valid choices.
- Enforcing constraints: It prevents the model from "thinking outside the options" and ensures only valid answers are generated.

Results

On Kaggle Leaderboard we got 0.469 score and highest current is 0.615.

Calculates using

Evaluation

Submissions are evaluated according to the Mean Average Precision @ 25 (MAP@25):

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

where U is the number of observations, $P(k)$ is the precision at cutoff k, n is the number predictions submitted per observation, and $rel(k)$ is an indicator function equaling 1 if the item at rank k is a relevant (correct) label, zero otherwise.

Once a correct label has been scored for *an observation*, that label is no longer considered relevant for that observation, and additional predictions of that label are skipped in the calculation. For example, if the correct label is `A` for an observation, the following predictions all score an average precision of `1.0`.

Demo Code Run Down

Looking Ahead?

- Trying hard to Incorporate Train data within Kaggle Computational Restrictions. So Relevant Knowledge not provided which can be efficient but computationally heavy.
- Looking forward to see responses of computationally heavy LLMs like Qwen 32B Preview and Open AI O1 Preview.
- Incorporation of Math heavy LLM Model by merging two LLM Models using Merge.kit from Arcee.AI.

Thank You

Questions ?