Final Project Options

Tatsuya Aoyama & Boyao Liu

Overview

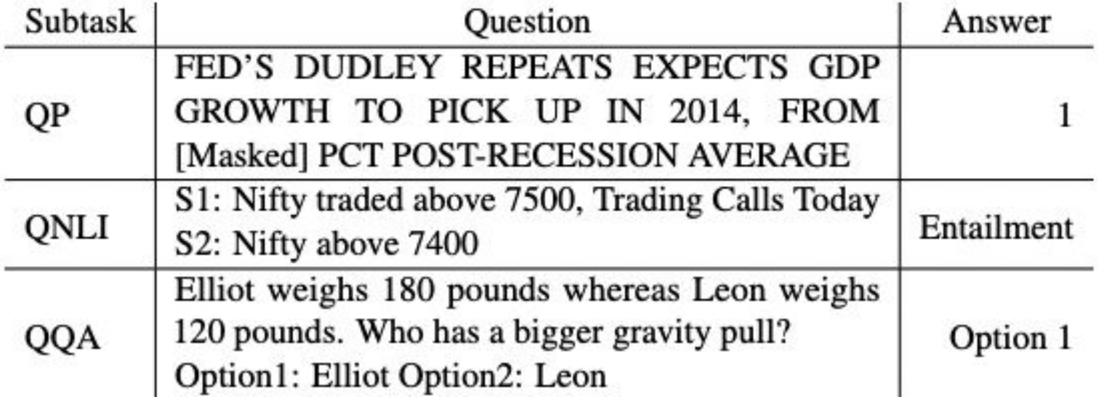
- Option 1: NumEval
- Option 2: BabyLM
- Option 3: Shroom
- Option 4: A Novel Task Defying Common Sense

Option 1: NumEval

https://sites.google.com/view/numeval/numeval?authuser=0

Option 1: NumEval

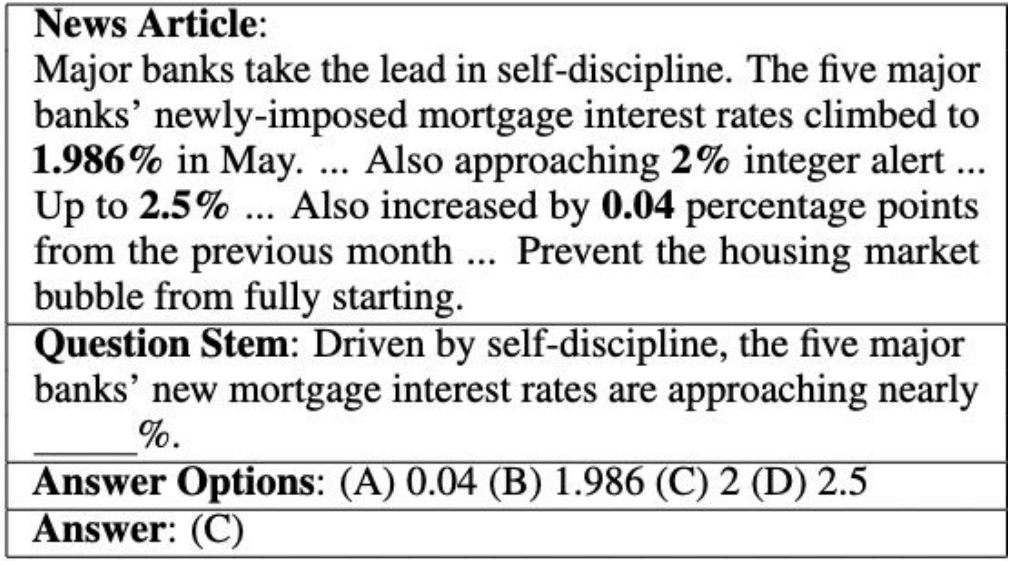
Task 1: Quantitative Understanding



Subtask	Question	Answer
QP	FED'S DUDLEY REPEATS EXPECTS GDP GROWTH TO PICK UP IN 2014, FROM [Masked] PCT POST-RECESSION AVERAGE	1
QNLI	S1: Nifty traded above 7500, Trading Calls Today S2: Nifty above 7400	Entailment
QQA	Elliot weighs 180 pounds whereas Leon weighs 120 pounds. Who has a bigger gravity pull? Option1: Elliot Option2: Leon	Option 1

Option 1: NumEval

Task 2: Reading Comprehension of Numerals (Chinese)



News Article:
Major banks take the lead in self-discipline. The five major banks' newly-imposed mortgage interest rates climbed to **1.986%** in May. ... Also approaching **2%** integer alert ... Up to **2.5%** ... Also increased by **0.04** percentage points from the previous month ... Prevent the housing market bubble from fully starting.

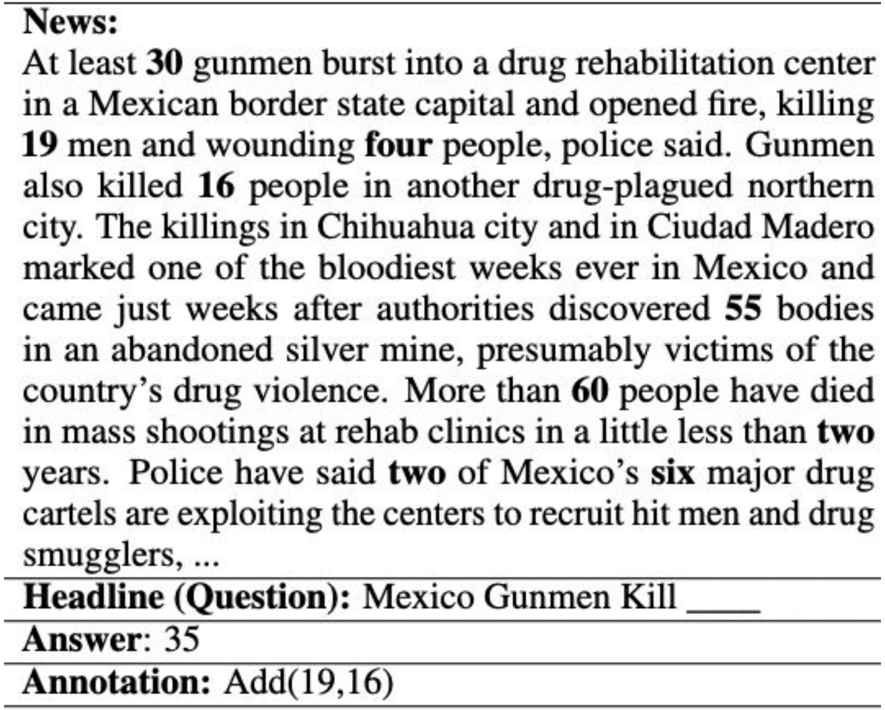
Question Stem: Driven by self-discipline, the five major banks' new mortgage interest rates are approaching nearly ______%.

Answer Options: (A) 0.04 (B) 1.986 (C) 2 (D) 2.5

Answer: (C)

Option 1: NumEval

Task 3: Numeral aware headline generation



News:
At least **30** gunmen burst into a drug rehabilitation center in a Mexican border state capital and opened fire, killing **19** men and wounding **four** people, police said. Gunmen also killed **16** people in another drug-plagued northern city. The killings in Chihuahua city and in Ciudad Madero marked one of the bloodiest weeks ever in Mexico and came just weeks after authorities discovered **55** bodies in an abandoned silver mine, presumably victims of the country's drug violence. More than **60** people have died in mass shootings at rehab clinics in a little less than **two** years. Police have said **two** of Mexico's **six** major drug cartels are exploiting the centers to recruit hit men and drug smugglers, ...

Headline (Question): Mexico Gunmen Kill ____

Answer: 35

Annotation: Add(19,16)

Option 2: BabyLM

https://babylm.github.io/index.html

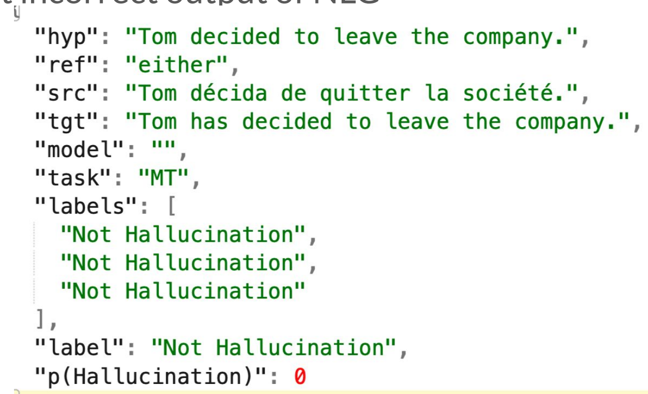
- Train an LM on 100M words or less
  - Strict: 10M or less words
  - Strict-small: 100M or less words
  - Loose: 100M or less + unlimited non-linguistic data + unlimited texts generated by the model itself
- Train set available, evaluation script available

Option 3: Shroom

https://helsinki-nlp.github.io/shroom/

Option 3: Shroom

- **Background**: NLG(Natural Language Generation) faces problem when generating contents, NLG model may generate contents look fluent but logically not suitable for original question. This phenomena is called **Hallucination**
- **Target**: Do binary classification on detecting fluent but incorrect output of NLG
- **Subtasks:**
  1. Definition Modeling
  2. Machine Translation
  3. Paraphrase Generation



```
{
  "hyp": "Tom decided to leave the company.",
  "ref": "either",
  "src": "Tom décida de quitter la société.",
  "tgt": "Tom has decided to leave the company.",
  "model": "",
  "task": "MT",
  "labels": [
    "Not Hallucination",
    "Not Hallucination",
    "Not Hallucination"
  ],
  "label": "Not Hallucination",
  "p(Hallucination)": 0
}
```

Piece of dev data for Track 2, machine translation

Option 3: Shroom

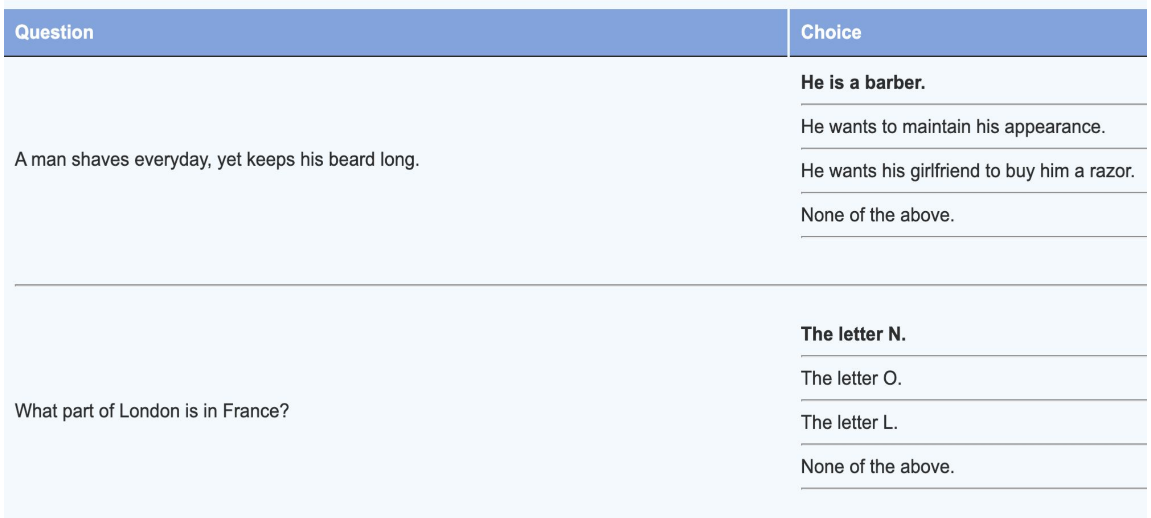
- Notable Pros/Cons for this option:
  - Built-in baseline kit and natural 50% baseline for binary classification
  - Their training data is claimed to be **not labeled**.
    - In their recommendation, they encourage to use some extra source to do labeling first
    - Another common solution is to do partition on dev/test set. However their labeled dev set is one thousandth the size of training data, which may not be ideal

Option 4: BRAINTEASER

https://brainteasersem.github.io/

Option 4: BRAINTEASER

- **Target**: Do multiple-choice Question Answering task, to observe subtle relationship between question and answer, defying preconceptions
- **Subtasks**:
  1. Sentence Puzzle
  2. Word Puzzle



Question	Choice
A man shaves everyday, yet keeps his beard long.	**He is a barber.**
	He wants to maintain his appearance.
	He wants his girlfriend to buy him a razor.
	None of the above.
What part of London is in France?	**The letter N.**
	The letter O.
	The letter L.
	None of the above.