BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY

(Autonomous Institute Affiliated to University of Mumbai)

MUNSHI NAGAR, ANDHERI (WEST), MUMBAI–400058

Department of Computer Engineering 2023-24

Natural Language Processing Mini Project

Title

"Comprehensive Answer Evaluation"

Bodhisatya Ghosh |

2021700026

Anish Gade | 2021700022

NLP LAB BATCH A

TE CSE DS | Batch B

Aim

To Evaluate Comprehensive Answers using sentence transformers.

Theory

To create a comprehensive answer evaluation system using Sentence Transformers and keyword matching, a multi-step approach can be employed. Firstly, input answers are encoded into high-dimensional vector representations using Sentence Transformers, capturing semantic nuances. Secondly, key concepts and keywords are extracted from the question using natural language processing techniques. These keywords serve as anchors for answer evaluation. Thirdly, similarity scores between answer vectors and keyword vectors are computed, determining relevance. Additionally, semantic similarity metrics between answer vectors and model answers can be calculated to gauge accuracy. Finally, aggregation of these scores enables holistic assessment, providing valuable insights into answer quality and relevance.

Introduction

First, we pass the answer and the answer from the synoptic in a sentence transformer. This transformer converts the sentences into embeddings that is the word vectors. We calculate the cosine similarity between the embeddings of both the answers. This score works as the factor for grading.

Approach 1 (Semantic Similarity using sentence transformers):

In this approach we used sentence transformers to create embeddings that is the word vectors of the synoptic and the input answer. We find cosine similarity between these answers that acts as a score factor.

Sentence Transformer Used:

We used an open-source sentence transformer form Hugging Face named "paraphrase-multilingual-MiniLM-L12-v2".

The metrics of the model are as follows:

SentenceTransformer(
  (0): Transformer({'max_seq_length': 128, 'do_lower_case': False}) with Transformer model: BertModel
  (1): Pooling({'word_embedding_dimension': 384, 'pooling_mode_cls_token': False, 'pooling_mode_mean_tokens': True, 'pooling_mode_max_tokens': False, 'pooling_mode_mean_sqrt_len_tokens': False})
)

Function used to calculate the score:

```
def get_similiarity(synoptic, answer):
    sentences =[synoptic, answer]
    model = SentenceTransformer('sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2')
    embeddings = model.encode(sentences)
    similiarity = cosine_similarity(embeddings[0].reshape(1,-1),embeddings[1].reshape(1,-1))

    return similiarity[0][0]
```

Approach 2 (Keywords Matching):

In this approach in synoptic the mentor can give a list of keywords. We add key-phrases to spacy keyword Matcher in forms of dictionaries. Then these keywords are matched with the words in the answer and then the we receive the score on the basis of number of words matched.

Function used:

```
def keyphrase_match_score(list_of_key_phrases: list[str], answer: str):
    nlp = spacy.load('en_core_web_sm')
    answer = clean_text(answer)
    score = 0

    for phrase in list_of_key_phrases:
        phrase = clean_text(phrase)

        m_tool = Matcher(nlp.vocab)

        vocab = [{'LOWER':word} for word in phrase.split()]
        m_tool.add('matcher',[vocab,])
        sentence = nlp(answer)
        if len(m_tool(sentence)) > 0:
            score+=1


    return score/len(list_of_key_phrases)
```

Output:

Sample question and answer

What are the main causes of climate change, and what can be done to mitigate its effects?

Synoptic: Climate change is primarily caused by human activities such as burning fossil fuels, deforestation, and industrial processes, which release greenhouse gases like carbon dioxide and methane into the atmosphere. These gases trap heat, leading to global warming and subsequent climate disruptions. To mitigate its effects, a multifaceted approach is necessary. This includes transitioning to renewable energy sources, implementing policies to reduce emissions, promoting afforestation and reforestation, adopting sustainable agricultural practices, enhancing energy efficiency, and encouraging international cooperation to set and achieve ambitious climate targets.

+ Code + Markdown

Given answer: Climate change is caused by various factors, including human activities such as burning fossil fuels and deforestation, which release greenhouse gases into the atmosphere. These gases trap heat, resulting in global warming and climate disruptions. To address climate change, efforts are needed to reduce emissions through transitioning to renewable energy sources, implementing policies to limit carbon emissions, and promoting sustainable practices in agriculture and industry.

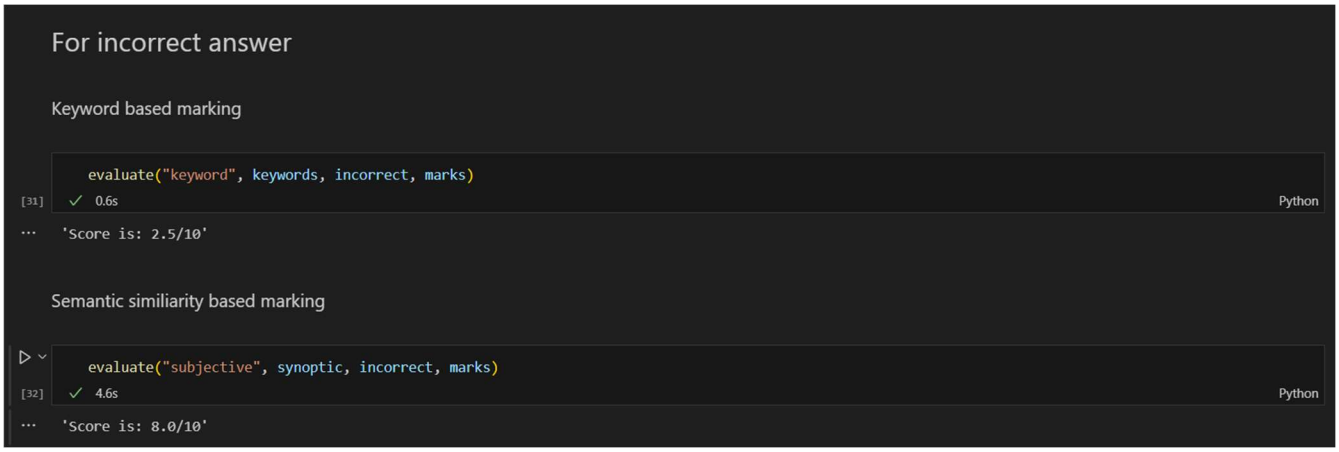
Incorrect keyword less answer: Climate change is mainly caused by natural fluctuations in the Earth's temperature and solar radiation. Human activities play a minor role, but they are not significant contributors to the overall climate change phenomenon. Therefore, there's little need for mitigation efforts beyond what nature can naturally adjust.

Correct keyword containing answer: Climate change is predominantly driven by human activities, such as the excessive use of fossil fuels and deforestation, resulting in the emission of greenhouse gases like carbon dioxide and methane. These gases trap heat in the atmosphere, leading to global warming and climate disruptions. To address this issue, it's imperative to implement comprehensive mitigation strategies. These include transitioning to renewable energy sources, enacting stringent policies to limit emissions, promoting afforestation and reforestation efforts, adopting sustainable practices across various sectors, and fostering international cooperation to tackle this global challenge effectively

List of sample keywords

- Climate change
- Causes
- Human activities
- Fossil fuels
- Greenhouse gases
- Mitigation
- Renewable energy
- Policies
- Afforestation
- Reforestation
- Sustainable practices
- International cooperation

```
synoptic = "Climate change is primarily caused by human activities such as burning fossil fuels, deforestation, and industrial processes, which release
imperfect = "Climate change is caused by various factors, including human activities such as burning fossil fuels and deforestation, which release gre
incorrect = "Climate change is mainly caused by natural fluctuations in the Earth's temperature and solar radiation. Human activities play a minor role
keyword_answer = "Climate change is predominantly driven by human activities, such as the excessive use of fossil fuels and deforestation, resulting in
keywords = [
    "Climate change",
    "Causes",
    "Human activities",
    "Fossil fuels",
    "Greenhouse gases",
    "Mitigation",
    "Renewable energy",
    "Policies",
    "Afforestation",
    "Reforestation",
    "Sustainable practices",
    "International cooperation"
```



Conclusion:

The Approach 1 turned out to be very effective as compared to approach 2. As approach 2 is very constrained. We will develop a weighted method to calculate score.

References

1]https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1184/reports/6907018.pdf

2] https://youtu.be/wFdFLWc-W4k?si=N12vYEt0HQevLiEw

3] fr2sql: Querying databases in French Benoit Couderc1Jeremy Ferrero2, 3 (1) Aix Marseille University, Marseille, France (2) Compilatio, 276 rue du Mont Blanc, 74540 Saint-Félix, France (3) LIG-GETALP, Grenoble Alpes University, France benoit.couderc@etu.univ-amu.fr , jeremy.ferrero@imag.fr