Text To MCQ Generation

Created by: Prashant Repale

Contents

1 Introduction

The generation of Multiple Choice Questions (MCQs) using Natural Language Processing (NLP) is an innovative approach that enhances educational assessment. By automatically creating questions from textual content, this technology aids in developing unbiased and diverse assessments, reducing the burden on educators. MCQ generation has applications in various settings, including academic testing, online learning platforms, and revision tools for students. It can facilitate personalized learning experiences, allowing learners to engage with material more interactively. As NLP models continue to advance, the potential for effective and adaptive MCQ generation will expand, transforming educational practices.

2 Workflow

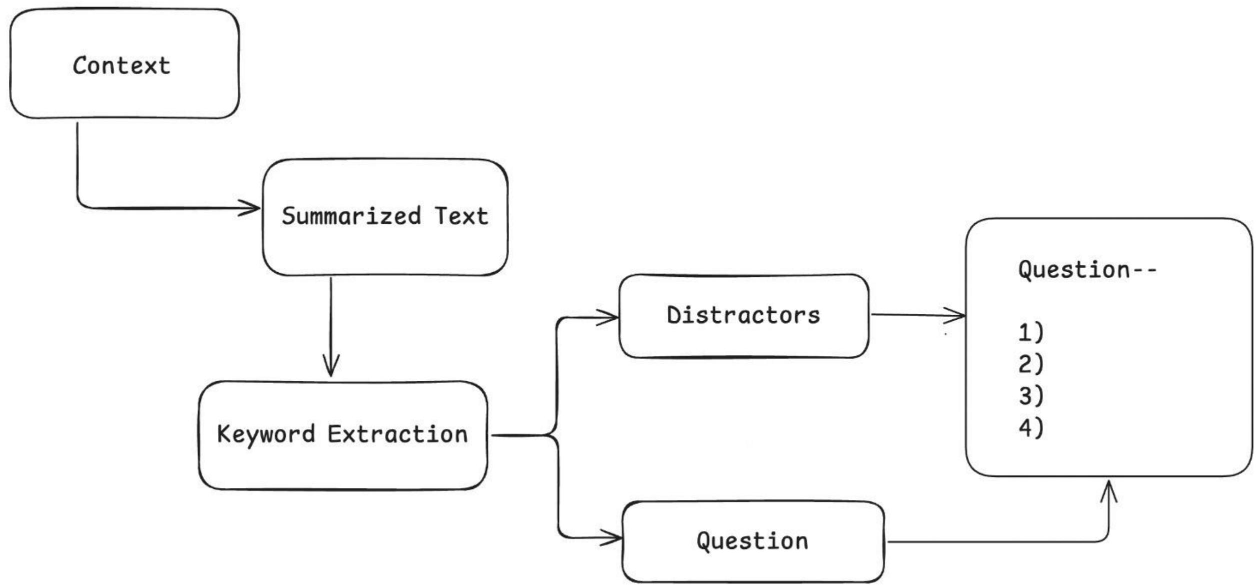

Figure 1: Project Workflow

3 Summarization

The T5 model, or *Text-to-Text Transfer Transformer*, represents a groundbreaking approach in natural language processing (NLP) by framing various tasks as text-totext problems. This innovative architecture unifies tasks such as translation, summarization, and classification under a single framework, streamlining the processes of model training and application.

T5 was pre-trained on a vast corpus of text using a masked language modeling objective. During this pre-training phase, input text is deliberately corrupted by replacing spans of tokens with a sentinel token, training the model to predict the missing tokens. This approach not only enhances the model's understanding of language but also equips it with the ability to handle diverse tasks effectively. For the specific task of summarization, T5 is fine-tuned on the XSum dataset, which comprises pairs of articles and their corresponding summaries. This fine-tuning process utilizes the same text-to-text objective, enabling the model to generate concise and coherent summaries that accurately capture the essence of the input articles.

4 KeyWord Extraction

In our project, we perform keyword extraction from summarized text using a systematic approach. First, we utilize the **nltk** library to download stopwords in English, which helps in filtering out common, less meaningful words from our text. We then employ the **pke** library's MultipartiteRank method to extract keywords from the original content.

The process involves loading the document and specifying parts of speech, focusing primarily on proper nouns (PROPN) and nouns (NOUN), as they are likely to carry the most significant meaning in the context. To refine our keyword candidates further, we exclude punctuation and common English stopwords. We then rank the candidates using the MultipartiteRank algorithm, which uses a graph-based random walk method to weigh the importance of terms.

Next, we use the **flashtext** library's **KeywordProcessor** to match the extracted keywords with those found in the summarized text. This helps identify the most relevant keywords that appear in both the original and summarized texts. Finally, we select and return the top four important keywords based on their presence in the summarized content.

This process ensures that the extracted keywords are both contextually relevant and concise, forming a foundation for generating meaningful multiple-choice questions later in the pipeline.

Example:
A lion lay asleep in the forest, his great head resting on his paws. The timid little mouse ran across the lion's nose and begged him to let him go. "please let me go and some day i will surely repay you," said the mouse. He was generous and finally let the mouse go; some days later, while stalking his prey, the lion was caught in an angry net.
Keyword exracted - ['lion', 'mouse', 'forest', 'net']

Distractor
In our project, the generation of distractors is a crucial step in creating effective multiple-choice questions. For this, we chose to use Sense2Vec, an advanced word embedding model, due to its superior ability to capture the different senses of a word. Unlike traditional models such as Word2Vec or GloVe, which can often return words that are contextually irrelevant, Sense2Vec associates words with their specific sense (e.g., 'apple—NOUN' vs. 'apple—ORG'). This means that when we look for distractors, the model understands whether we are referring to "apple" as a fruit or as a company, resulting in much more contextually accurate suggestions.
We start by extracting similar words using Sense2Vec and apply our custom filter to ensure that all the suggested distractors share the same sense as the original word. This step is vital because we only want distractors that are semantically aligned with the target word, ensuring that they make sense in the context of the question. For instance, if the original word is "bank—NOUN" (meaning a financial institution), we don't want distractors like "riverbank" or "shore", which would lead to confusion. Next, we incorporate the Maximal Marginal Relevance (MMR) algorithm to fine-tune the list of distractors. The MMR algorithm plays an important role by ensuring that we strike a balance between two key objectives: relevance and diversity. Here's why these two factors are critical:

- **Relevance**: The distractors should be closely related to the question and the context. To achieve this, we compute the similarity between each distractor and the original sentence using sentence embeddings from the 'SentenceTransformer' model. This ensures that all the selected distractors are contextually appropriate and make sense when compared to the original content.

- **Diversity**: In addition to relevance, we also want to avoid having distractors that are too similar to one another, as this would make the question less challenging for the test taker. The MMR algorithm helps us minimize the similarity between the distractors themselves. It does this by selecting distractors that are both related to the original word and distinct from one another, leading to a more varied and engaging set of options for the MCQ.

The MMR process starts by identifying the most relevant word first and then gradually adds more diverse words based on their similarity to the initial selection. By doing this, we ensure that each distractor not only fits the context but also introduces subtle distinctions that make the question more thought-provoking.
In addition, we use a similarity threshold to ensure that distractors aren't too close in meaning to the correct answer, making the question appropriately challenging. The SentenceTransformer model helps us calculate cosine similarity between the embeddings of both the original sentence and the distractors, and MMR ensures that the final selection strikes the right balance between relevance and variety. Ultimately, the combination of Sense2Vec and the MMR algorithm allows us to generate high-quality distractors that are contextually accurate, diverse, and challenging, contributing significantly to the effectiveness of our text-to-MCQ generation process. This approach ensures that the MCQs we generate are more sophisticated and better suited to evaluating the knowledge of the test taker.

Example:
Barack Obama: John McCain,George W Bush,Sarah Palin,Bill Clinton *India:* Brunei,Nigeria,Singapore,Nepal

5 Question Generation

LLaMA (Large Language Model Meta AI) 3.1 is a state-of-the-art transformer-based language model specifically designed to excel in various instruction-based tasks. For this project, we fine-tuned the 8 billion parameter variant, LLaMA 3.1-8b Instruct, to generate questions based on a given answer and its relevant context.
To effectively fine-tune the model, we defined the following components:

- C: A paragraph or context of text.
- A: An answer extracted from C.
- Q: A human-generated question that, when posed within C, would elicit A as the answer.

We utilized the SQuAD dataset, where C and A served as inputs and Q as the target output. The fine-tuning of the LLaMA model was carried out using the SFTTrainer class.

The model underwent fine-tuning for 2 epochs, employing an 8-bit variant of the paged adamw optimizer, which is an adaptive optimizer that incorporates weight decay. The learning rate was set to 1e-4 to ensure effective convergence.

Throughout the fine-tuning process, evaluations were performed at regular intervals to monitor the model's performance and mitigate the risk of overfitting. This rigorous evaluation process allowed us to refine the model's capabilities and enhance its question-generation performance.

Model Performance

Example from the fine-tuned Model:

> **Context:** The human body has 206 bones, which provide structure, protect vital organs, and facilitate movement. The largest bone is the femur, located in the thigh, while the smallest is the stapes, found in the middle ear. Maintaining bone health is essential, as they can weaken with age or due to certain medical conditions.
>
> **Answer:** 206

The model generates the following question:
Generated Question: How many bones are in the human body?

6 Results

Example:

Text:

Elon Musk completed his acquisition of Twitter in October 2022; Musk acted as CEO of Twitter until June 2023 when he was succeeded by Linda Yaccarino. Under Musk, Twitter was rebranded to X on July 23, 2023, and its URL changed from twitter.com to x.com on May 15, 2024.
X is one of the top social media platforms and the fifth-most-visited website in the world as of June 2024. Users can share posts containing text messages, images, and videos and interact with other users' content through likes and reposts. X offers additional features such as direct messaging, video and audio calling, bookmarks, lists, communities, a chatbot, and the social audio feature Spaces.
Founded in March 2006 by Jack Dorsey, Noah Glass, Biz Stone, and Evan Williams as Twitter, it underwent a rebranding in July 2023 after being acquired by Elon Musk in 2022. Now operating as X, the platform closely resembles its predecessor but includes additional features such as long-form texts, account monetization options, audio-video calls, integration with xAI's Grok chatbot, job search, and a verification process accessible to premium users. Several Twitter legacy features were removed from the site after Musk acquired Twitter, including Circles, NFT profile pictures, and pronouns in profiles. Musk aims to transform X into an "everything app", akin to WeChat.

Generated Multiple Choice Questions:

Question: Who acquired Twitter in 2022?
Ans: Elon Musk
Warren Buffet
Jeff Bezos
Steve Jobs

Question: When was X the fifth most visited website?
Ans: June
October
January March

The proposed text-to-MCQ system follows a structured workflow designed to automatically generate multiple-choice questions. Initially, the input context is summarized to extract key information. From this summarized text, important keywords are identified through keyword extraction techniques, which serve as the basis for generating the question. Simultaneously, distractors are generated to accompany the correct answer, ensuring plausible alternatives for the MCQ. The integration of both distractors and the correct answer results in a well-structured multiple-choice question, optimizing the process of question generation for educational assessments.