ML REPORT

PARA-TO-QUESTION GENERATOR

Machine Learning Project

UML501

Department of Computer Science and Engineering

Submitted to:
Dr. Raman Kumar Goyal

Submitted by:
Vansh Kansal (102203021)
Priyanshu Naval
(102203063)

INTRODUCTION:

The Text-to-Question Generator is designed to transform paragraphs into meaningful and relevant questions using advanced natural language processing techniques. This project employs a fine-tuned T5 (Text-to-Text Transfer Transformer) model trained on the SQuAD 1.1 dataset to generate high-quality questions from user-provided text inputs.

With the growing demand for automated educational tools, this system is valuable for educators, content creators, and e-learning platforms, enabling efficient testing of comprehension and learning outcomes.

TECHNOLOGY STACK:

This project uses state-of-the-art machine learning frameworks for text generation:

- T5 Transformer Model: A pretrained transformer-based model optimized for text-to-text tasks, used for converting paragraphs into questions.
- PyTorch: The core deep learning framework employed for model fine-tuning and inference.
- PyTorch Lightning: Simplifies training and ensures scalability.
- Frontend: Developed with HTML, CSS (Tailwind), providing a user-friendly interface.
- Flask: A lightweight web framework used to create a simple web application for users to upload images and view the extracted calorie values.

WORKFLOW OF THE SYSTEM:

The system follows a structured pipeline for question generation:

1. Text Input:
    - Users input a paragraph via the interactive web interface.
2. Text-to-Question Model:
    - The input text is processed by a fine-tuned T5-small model.
    - The model generates relevant questions by analyzing the context and extracting essential information.
3. Result Display:
    - Generated questions are displayed to the user in real time.
4. Customization Options:
    - Users can modify parameters such as the number of questions generated and the complexity level.

DATASET:

The SQuAD (Stanford Question Answering Dataset) 1.1 was used to fine-tune the T5-small model.

1. Dataset Features:
    - Over 100,000 question-answer pairs.
    - Data sourced from a wide range of topics including history, science, and current events.
    - Emphasizes intensive reasoning and contextual understanding.
2. Training Details:
    - Model trained for 4 epochs with the following hyperparameters:
        - Source Max Token Length: 512
        - Target Max Token Length: 128
        - Batch Size: 8
        - Learning Rate: 0.00005
    - To prevent overfitting, the answer tokens were randomly replaced with a [MASK] token during training.

CONCLUSION:

The Para-to-Question Generator demonstrates the potential of transformer-based models in automating educational tasks. By fine-tuning T5 on SQuAD 1.1, the system efficiently generates questions, enabling users to test comprehension in real time.

This project highlights the application of cutting-edge NLP techniques to solve real-world problems, offering accuracy, scalability, and user-friendly interaction.

FUTURE ENHANCEMENTS:

1. Domain-Specific Fine-Tuning: Customize the model for specific subjects like mathematics or Sciences.

2. Multilingual Support: Extend capabilities to support languages other than English.

3. Additional Question Types: Enable generation of multiple-choice, true/false, and descriptive questions.

4. Performance Optimization: Improve response time for real-time usage.

5. API Development: Provide endpoints for integration with external systems or applications.