Individual Final Report

NLP 2023 Fall

HaeLee Kim

1. Introduction

In the dynamic landscape of news consumption, our team project endeavors to create a Smart News Reader empowered by advanced Natural Language Processing (NLP). This tool transcends traditional news consumption by incorporating features like fetching, summarization, zero-shot classification, keyword extraction, and conversational question-answering.

My individual contributions within this collaborative framework were focused on summarization, zero-shot classification, and many-to-one translation. As part of a cohesive team, each member's unique expertise contributed to the realization of our project's overarching goals.

2. Description of Individual Work

In my individual work, I concentrated on models for three components: Summarization, Zero-Shot Classification, and Translation.

2-1. Summarization:

I will elaborate on the BART-Large-CNN model's implementation, addressing challenges associated with the model's token limitations. It's essential to note that the LangChain part to improve the model accuracy was handled by another teammate, Arjun Bingly.

> 2.2 Summarization The Summarization component plays a pivotal role in distilling extensive news articles into succinct and informative summaries. Leveraging advanced natural language processing techniques, the Summarizer class utilizes the BART-Large-CNN model. This model, pre-trained on the CNN news Dataset, since this was the most extensive news article dataset available, we did not have to fine-tune the model.

2-2. Zero-Shot Classification:

I will delve into the integration of the BART-Large-MNLI model into the ZeroShotClassifier class, approaching to categorizing news articles without specific training data.

> 2.3 Zero-Shot Classification The main idea behind this module is that the users will be able to check if a particular news article is about the a particular personalized label category. The Zero-Shot Classification module addresses the challenge of categorizing news articles without the need for specific training data for each category. Leveraging the BART-Large-MNLI model fine-tuned on CNN (AyoubChLin/Bart-MNLI-CNN news), integrated into the ZeroShotClassifier class, this module excels in assigning relevant labels to news articles.

2-3. Translation: Failed to Implement:

I tried to use Many-to-One translation module using the MBART model, discussing the encountered challenges and the ultimate setback in implementation.

2.6 Translation - Failed to Implement The Translation component ensures the accessibility of news articles across diverse linguistic audiences. The Many-to-One translation module, equipped with the MBART model, is proficient in translating news articles from various source languages to English. The NewsTranslator class dynamically handles language codes, offering an efficient and flexible translation mechanism.

3. Code Contribution Percentage – 30%