GROUP-20 ; Efficient Summarization of Healthcare Responses

Dhairya
Roll No: 2022157
dhairya22157@iiitd.ac.in

Harsh Vishwakarma
Roll No: 2022205
harsh22205@iiitd.ac.in

Pandillapelly Harshvardhini
Roll No: 2022345
pandillapelly22345@iiitd.ac.in

Abstract

Today, Natural Language Processing has become very advanced which has enabled the development of large and complex models which are capable of generating high quality summaries from the given text taken from the Internet for its corresponding task. In this work, we explore the effectiveness of FLAN-T5 and PLASMA models for perspective-aware healthcare summarization tasks. Meanwhile the existing methods only focuses on general purpose summarization, they often fail to capture the contextual and perspective based that are critical for real world applications. Our study tells how these models handle multi-perspective summarization by incorporating structured prompt-based learning and energy controlled loss function. We evaluate there performance using the evaluation metrices like BLUE, BERTscore etc. highlighting their effectiveness in generating concise and contextually related summaries.

1 Introduction

With the rise of digital platforms such as Quora, Reddit, and Yahoo!—community question-answering (CQA) forums—how people seek and share information has been revolutionized. These platforms allow users to discuss various topics, such as healthcare, technology, and personal experiences, enabling the collaborative exchange of knowledge.

Online platforms have become a popular space for medical discussions, with communities such as r/AskDocs, r/DiagnoseMe, r/Medical_Advice, and r/Medical serving as hubs for users seeking health-related guidance. These forums cater to a wide range of inquiries, from general health concerns and symptom assessments to questions about medical procedures and professional practices.

While these platforms provide valuable support and collective knowledge, the vast number of responses—ranging from medically accurate insights to personal anecdotes and subjective opinions—can make it difficult for users to extract reliable and meaningful information. This challenge is further amplified by the variation in response quality, the lack of standardized medical verification, and the potential for misinformation, making it crucial for users to critically assess the advice they receive.
Summarizing these responses in structured and perspective aware manner is important for improving the accessibility and comprehension. Traditional methods of summarization focuses on extracting the key information, they often overlook the importance of capturing the multiple viewpoints from the text. Recent advancements in transformers based models, particularly, FlanT5 and PLASMA, these models have addressed the problem of traditional model as mentioned above by integrating the prompt based learning. In this paper, we explore the efficacy of these models in generating structured, perspective-driven summaries and assess their performance using datasets and evaluation metrics.

2 Dataset

The dataset consists of three primary files: train.json, valid.json, and test.json, which are structured to support perspective-aware summarization tasks. A dataset includes: a unique identifier (uri) for question-answer threads, a user-posted question, optional context, a set of answers, labeled answer spans categorizing responses into perspectives (Information, Cause, Suggestion, Experience), and labeled summaries providing concise perspective-specific insights.

3 Problem Statement

In the digital age, online health forums and medical Q&A platforms serve as vital resources for individuals seeking medical advice. However, the sheer volume of responses—ranging from expert-

backed insights to personal experiences—poses a significant challenge in extracting reliable and meaningful information. The lack of standardized verification, varying response quality, and potential for misinformation further complicate the process, leaving users uncertain about which advice to trust.

Our goal is to develop a platform that generates summaries by extracting diverse perspectives—such as causes, suggestions, and personal experiences—from multiple responses to a single medical question. By utilizing advanced recommendation algorithms, our system will intelligently generalize user-posed medical queries and curate the most accurate, relevant, and evidence-based information. It will filter out inconsistencies, prioritize medically verified content, and present users with a well-rounded, credible, and informative summary, ensuring improved access to trustworthy health information.

4 High-Level Plan

To improve the effectiveness of perspective based summerization, we will do an iterative refinement of our approach by following steps:

- EDA and Feature Engineering: performing Exploratory data analysis and feature engineering on the dataset.
- Expanding model variants: Exploring models beyond FLAN-T5 and PLASMA which will improve the summerization efficiency.
- Fine-Tuning Strategies: Try to apply advanced fine tuning techniques like LoRA and instruction tuning to enhance the model perspective based summarization based ability.
- Loss function optimization: Exploring other optimization techniques, such as contrastive learning, to improve alignment between between generated summaries and human-annotated references.
- Evaluation Metrices expansion: Exploring additional evaluation metrics which will help us to assess better model performance.
- Basic command line interface / User interface, where user can generate summaries by giving prompt.

5 Approach

For now, we have decided on the following approach, which may be adjusted in the future as needed.

- Analyzing the dataset to understand response patterns, variations and perspective distributions while extracting linguistic and semantic features for perspective alignment.
- Perspective Classification Enhancement: Replacing RoBERTa model with alternative classifiers like SpanBERT or etc. to better distinguish between perspective categories.
- Summarization Model Optimization: Replacing or fine-tuning FLAN-T5 and PLASMA with more recent models, like Mistral-7B, which have demonstrated superior long-context understanding.
- Hyperparameter Tuning: Conducting extensive tuning experiments to identify optimal training configurations for improved accuracy and efficiency.
- Replacing the energy-based loss function with contrastive learning.

References

[1] No perspective, no perception!! Perspective-aware Healthcare Answer Summarization Available at: arxiv.org

[2] FLAN-T5, HuggingFace Available at: Link

[3] SpanBERT Link