Project Title

Perspective-aware Summarization for Healthcare QA

Brief Description of the Problem Statement

Healthcare Community Question-Answering (CQA) forums contain diverse, often off-topic answers, making it challenging for users to extract useful information. Existing summarization methods ignore varied perspectives (e.g., cause, suggestion, experience) in responses. This project addresses the need for perspective-aware summaries to improve information accessibility.

Brief Description of the Dataset Used

The PUMA dataset, comprising 3,167 healthcare CQA threads and over 6,000 perspective-specific summaries. Data splits include:

- **train.json**: 2,236 samples
- **valid.json**: 959 samples
- **test.json**: 640 samples
  Each thread includes annotated answer spans and summaries for perspectives like *Information*, *Suggestion*, *Experience*, *Question*, and *Cause*.

Baseline 1 Model Description

The model is a fine-tuned flan-t5-small on the "INFORMATION_SUMMARY" task. Inputs combine questions, answers, and a perspective prompt, with training parameters: 3 epochs, batch size 4, learning rate 3e-4, and FP16 acceleration.

Baseline 1 Results on the Relevant Evaluation Metric

- **BLEU Score:** 0.0932 (validation), 0.0942 (test).
- **BERTScore F1:** ~0.8838 (validation), ~0.8833 (test).
  Results indicate moderate semantic alignment (BERTScore) but low n-gram overlap (BLEU), typical for abstractive summarization.

Baseline 2 Model Description

Uses FLAN-T5-small fine-tuned for summarization, trained on question-answer pairs to generate "INFORMATION_SUMMARY" outputs (25 epochs, batch size 8).

Baseline 2 Results on the Relevant Evaluation Metric

- **BLEU Score:** 0.24 (validation), 0.22 (test).
- **BERTScore F1:** ~0.58 (validation), ~0.57 (test).
  Results indicate moderate semantic alignment with reference summaries but limited n-gram overlap, suggesting generated summaries capture general meaning but lack precise phrasing.

Final Model Description

The final model is a fine-tuned BART-base (facebook/bert-base) transformer model, trained end-to-end on the PUMA dataset for multi-perspective healthcare answer summarization. It uses standard fine-tuning with cross-entropy loss, AdamW optimizer, 3-5 epochs, and a learning rate of 5*10^(-5) , eliminating the need for handcrafted prompts or auxiliary losses. The model generates summaries tailored to perspectives like Information, Suggestion, Experience, Question, and Cause.

Final Model Results on the Relevant Evaluation Metric

- **BLEU:** 0.094
- **BERTScore F1:** 0.894

  These results demonstrate strong semantic alignment (high BERTScore) with reference summaries, outperforming Flan-T5 baselines in semantic metrics despite lower n-gram overlap (BLEU), typical for abstractive summarization tasks.