

HEALTHCARE ANSWER SUMMARIZATION

GROUP 9



INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

RIYA GUPTA (2022410)
SWARA PAREKH (2022524)
VIMANSH MAHAJAN (2022572)

NLP FINAL PROJECT



Problem Statement & Motivation



- **Problem:**

- Healthcare answers online are often long and hard to digest.
- Users struggle to extract key perspectives like causes, suggestions, etc.

- **Goal:**

- Build an NLP system that provides **concise, perspective-aware summaries**.

- **Impact:**

- Supports both healthcare professionals and the general public.
- Enhances clarity and accessibility.



- **Perspective-aware Healthcare Answer Summarization: [Link](#)**
 - Proposed PLASMA, a Flan-T5-based summarization model using prefix tuning and an energy-controlled loss function that integrates perspective, tone, and anchor information. It outperformed strong baselines (e.g., BART, GPT-2) with top scores: ROUGE-L F1 of 21.38, BERTScore of 0.869, and BLEU of 0.0405.
- **Comparative Analyses of Transformer Models for Text-Based Emotion Recognition: [Link](#)**
 - The paper compares four transformer models—BERT, RoBERTa, DistilBERT, and XLNet—for emotion recognition on the ISEAR dataset, classifying text into seven emotions: Anger, Disgust, Sadness, Shame, Fear, Joy, and Guilt. RoBERTa achieved the best performance among them.

- **Classifier:** Two-heads - **Span Tagging** (Performs BIO tagging to highlight text spans corresponding to each perspective) and **Multi-label Classification** (Predicts presence of 5 perspectives — Information, Suggestion, Experience, Question, Cause)
- **Generator:** Uses “facebook/bart-large-cnn” fine-tuned with LoRA for efficient training. Uses the perspectives generated by the classifier, along with the Question, Context and Answers as input, to output one Summary per perspective.
- **Hard Examples Fine Tuning:** Extract examples from train dataset that have BERTScore < 0.84 and fine-tune the model.

- **PIPELINE :**

Classifier (Perspectives) → Generator (Summaries) → Evaluate (BLEU and BERTScore)

Dataset & Baselines



Dataset:

- Modified PUMA dataset: 3,835 threads
- Split: Train (2,236), Val (959), Test (640)

Changes from original PUMA:

- Removed “Treatment” perspective
- Merged “Clarification” into “Information”
- Final Perspectives : Information, Suggestion, Experience, Question, Cause

Baselines:

- Flan-T5 + LoRA
- GPT-OPT (GOT) + LoRA

Results & Comparison with Baselines



Perspective	Bart + LoRA BLEU	Bart + LoRA BERTScore	FlanT5	GOT
Information	10.44	0.8770	4.16 / 0.8491	4.54 / 0.8324
Suggestion	6.43	0.8635	2.42 / 0.8364	2.51 / 0.8223
Experience	3.92	0.8465	1.96 / 0.8369	1.78 / 0.8179
Question	0.51	0.8376	0.80 / 0.8366	0.45 / 0.8063
Cause	6.74	0.8676	2.51 / 0.8544	1.74 / 0.8215
Average	5.61	0.8782	2.37 / 0.8427	2.21 / 0.8201

(BLEU/ BERTScore)

(BLEU/ BERTScore)

Results of the Pipeline Model



Perspective	Precision	Recall	F1-score	Support
INFORMATION	0.875	0.947	0.910	735
SUGGESTION	0.894	0.934	0.914	595
EXPERIENCE	0.850	0.899	0.874	316
QUESTION	0.855	0.461	0.599	102
CAUSE	0.650	0.561	0.602	139

Evaluation Metrics of the Classifier Model

Table 1: Results of the Classifier Model

Evaluation Metrics of the final Pipeline Model

Perspective	BLEU	BERTScore
INFORMATION	10.4398	0.8770
SUGGESTION	6.4325	0.8635
EXPERIENCE	3.9180	0.8465
QUESTION	0.5080	0.8376
CAUSE	6.7420	0.8676
Average	5.6081	0.8782

Table 4: Perspective-wise BLEU and BERTScore results

Why This Approach Worked



- **BART** - pretrained as both a denoising autoencoder and a sequence-to-sequence model. Captures long-range dependencies and generating coherent summaries.
- **LoRA** - efficient fine-tuning by introducing lightweight parameter updates, leading to faster convergence and better retention of pretrained knowledge.
- **Hard examples** - model focus on challenging and underrepresented cases, thereby improving robustness and promoting a deeper understanding of context.

Conclusion & Future Work



- **Conclusion:**

- Introduced a scalable, perspective-aware summarization pipeline.
- Performance improvement over existing methods.

- **Future Directions:**

- Joint training of classifier and generator.
- Advanced hard example mining (contrastive/uncertainty-based).
- User-controllable summarization interface.